

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 20, 2004, 19:11:02 ; Search time 64 Seconds  
(without alignments)  
7753.307 Million cell updates/sec

Title: US-09-647-661-1  
Perfect score: 1559  
Sequence: 1 tgcagattttacagagc.....gggtgttttcagcggttc 900

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1124875 seqs, 275673149 residues

Total number of hits satisfying chosen parameters: 2249750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USPO1 spool/p/US09647661/runat\_20042004\_135346\_2594/app\_query.fasta\_1.1095  
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum2  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09647661@cgn 1 1.13 @runat\_20042004\_135346\_2594  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXI=7

Database : Published Applications AA:

1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pcp.\*  
2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pcp.\*  
3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pcp.\*  
4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pcp.\*  
6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pcp.\*  
7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pcp.\*  
8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pcp.\*  
9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pcp.\*  
10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pcp.\*  
11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pcp.\*  
12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pcp.\*  
13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pcp.\*  
14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pcp.\*  
15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pcp.\*  
16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pcp.\*  
17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pcp.\*  
18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

1	1072	68.8	210	10	US-09-882-227-180
2	1067	68.4	210	15	US-10-364-397-29
3	1041	66.8	210	12	US-10-335-977-6487
4	636	40.8	134	12	US-10-335-977-6486
5	626	40.2	127	12	US-10-335-977-6485
6	448	28.7	97	12	US-10-335-977-6484
7	427	27.4	91	12	US-10-335-977-6483
8	246	15.8	284	9	US-09-895-913A-102
9	229	14.7	283	12	US-10-335-977-5526
10	218.5	14.0	220	15	US-10-364-397-17
11	198.5	12.7	201	15	US-10-364-397-23
12	190	12.2	217	12	US-10-335-977-6413
13	185.5	11.9	202	15	US-10-364-397-12
14	182	11.7	221	14	US-10-169-048-2
15	174	11.2	200	15	US-10-364-397-19
16	172.5	11.1	209	15	US-10-364-397-10
17	153	9.8	174	15	US-10-364-397-21
18	150	9.6	198	10	US-09-769-744A-30
19	136.5	8.8	205	15	US-10-364-397-18
20	124	8.0	172	15	US-10-364-397-20
21	116	7.4	185	12	US-10-335-977-6272
22	114	7.3	197	12	US-10-282-122A-53487
23	113	7.2	205	14	US-10-156-761-8103
24	106.5	6.8	79	12	US-10-335-977-6411
25	105.5	6.8	219	12	US-10-282-122A-61643
26	102.5	6.6	289	12	US-10-236-417-90
27	97.5	6.3	90	12	US-10-335-977-6412
28	93.5	6.0	235	14	US-10-153-362-7
29	92.5	5.9	218	12	US-10-282-122A-53372
30	92	5.9	234	9	US-09-800-729-150
31	92	5.9	234	11	US-09-833-245-2210
32	91.5	5.9	241	14	US-10-221-945-1
33	91	5.8	1125	12	US-10-282-122A-47213
34	90.5	5.8	214	14	US-10-239-567-11
35	89.5	5.7	215	15	US-10-307-724-122
36	89	5.7	1001	12	US-10-335-977-5785
37	89	5.7	1001	12	US-10-335-977-5786
38	88.5	5.7	141	14	US-10-153-382-23
39	88.5	5.7	214	12	US-10-411-037-55
40	88.5	5.7	214	12	US-10-411-026-55
41	88.5	5.7	237	9	US-09-056-160B-100
42	88.5	5.7	237	14	US-10-234-671-100
43	88.5	5.7	491	13	US-10-011-125-2
44	88	5.6	625	12	US-10-282-122A-46895
45	87.5	5.6	214	12	US-10-364-953-1

#### ALIGNMENTS

#### RESULT 1

US-09-882-227-180  
Sequence 180, Application US/09882227  
Publication No. US20030158396A1  
GENERAL INFORMATION:  
APPLICANT: Kleathous, Harold  
APPLICANT: Al-Garawi, Amal  
APPLICANT: Miller, Charles  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Oomen, Raymond P.  
TITLE OF INVENTION: Identification of Polynucleotides  
TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in th  
TITLE OF INVENTION: Genome  
FILE REFERENCE: 06132/047002  
CURRENT APPLICATION NUMBER: US/09/882,227  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 08/902,615  
PRIOR FILING DATE: 1997-07-29  
NUMBER OF SEQ ID NOS: 638  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 180  
LENGTH: 210  
TYPE: PRT  
ORGANISM: Helicobacter pylori

US-09-882-227-180

Alignment Scores:

Pred. No.: 1 59e-110 Length: 210  
Score: 1072.00 Matches: 207  
Percent Similarity: 99.05% Conservative: 1  
Best Local Similarity: 98.57% Mismatches: 2  
Query Match: 68.76% Indels: 0  
DB: 10 Gaps: 0

US-09-647-661-1 (1-900) x US-09-882-227-180 (1-210)

QY 151 ATGAATTTTGGATCAGGAAAAAGAGCACTATTAAACGAGCGCCATTCTTCAAG 210  
Db 1 MetLysPheLeuAspGlnGluYsArgGlnLeuLeuAsnGluArgHisSerCysLys 20  
QY 211 ATGTTTCACAGCCATTATGAGTTTCTAGTGAAGATTAGAAGAAATCGTGAATCGCC 270  
Db 21 MetPheAspSerHisTyrGluPheSerThrGluLeuGluGluLeuAlaGluLeuAla 40  
QY 271 AGGCTATCGCAAGCTCTTACACACGAGCCATGCGATTCTTGTAGTGTACTATAAG 330  
Db 41 ArgLeuSerProSerSerTyrAsnThrGlnProTrpHisPheValMetValThrAspLys 60  
QY 331 GATTTAATAAACAATAATGAGTCAGTCACAGCTACTTTAATCAAGAAATGATTAAACCGCT 390  
Db 61 AspLeuLysGlnIleAlaAlaHisSerTyrPheAsnGluLeuMetIleLysSerAla 80  
QY 391 TCAGCGTTAATGTTGATGCTCTTTAAGACCTAGCGAGTTGTACACACGCGCCATTAC 450  
Db 81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100  
QY 451 ATGCAAAACCTTTACCGGAGTCTTATAGGTAGAGTATCCCTCTTTTGTCTCAATG 510  
Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120  
QY 511 CTTCGCGTGAGATTCAACACGAGCATGCAAGAGATTAGAAAGCTATATTAGAGCAATGC 570  
Db 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGlnCys 140  
QY 571 TATATCGCTGTGGGCAAAATTTGATGGCGTGTAGCTTAATGGATGTAGTGTGCATT 630  
Db 141 TyrIleAlaValGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
QY 631 ATTGAGGCTTTGATCCTTTAAAGTGGTGAAGTTTGAAGAGCGCTATCAATAAGCT 690  
Db 161 IleGlyGlyPheAspProLeuLysValGlyValLeuLeuGluArgIleAsnLysPro 180  
QY 691 AAAATCGCATGCTTGTATCGCTTTGGCAAGAGGTGGCAGAGCGCCAAAATCAAGA 750  
Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
QY 751 AAATCAAAAGTTGATGCGATTCTTGGTTG 780  
Db 201 LysSerLysValAspAlaIleThrIlePhe 210

RESULT 2

US-10-364-397-29  
; Sequence 29, Application US/10364397  
; Publication No. US20040014191A1  
; GENERAL INFORMATION:  
; APPLICANT: Minton, Nigel  
; APPLICANT: Anlezark, Gail  
; APPLICANT: Vaughan, Thomas  
; TITLE OF INVENTION: Microductase Enzymes  
; FILE REFERENCE: 1581/0650001  
; CURRENT APPLICATION NUMBER: US/10/364,397  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: 09/913,068  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: PCT/GB00/00431  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: UK 9903019.9

US-09-647-661-1 (1-900) x US-10-364-397-29 (1-210)

QY 151 ATGAATTTTGGATCAGGAAAAAGAGCACTATTAAACGAGCGCCATTCTTCAAG 210  
Db 1 MetLysPheLeuAspGlnGluYsArgGlnLeuLeuAsnGluArgHisSerCysLys 20  
QY 211 ATGTTTCACAGCCATTATGAGTTTCTAGTGAAGATTAGAAGAAATCGTGAATCGCC 270  
Db 21 MetPheAspSerHisTyrGluPheSerThrGluLeuGluGluLeuAlaGluLeuAla 40  
QY 271 AGGCTATCGCAAGCTCTTACACACGAGCCATGCGATTCTTGTAGTGTACTATAAG 330  
Db 41 ArgLeuSerProSerSerTyrAsnThrGlnProTrpHisPheValMetValThrAsnLys 60  
QY 331 GATTTAATAAACAATAATGAGTCAGTCACAGCTACTTTAATCAAGAAATGATTAAACCGCT 390  
Db 61 AspLeuLysGlnIleAlaAlaHisSerTyrPheAsnGluLeuMetIleLysSerAla 80  
QY 391 TCAGCGTTAATGTTGATGCTCTTTAAGACCTAGCGAGTTGTACACACGCGCCATTAC 450  
Db 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProThrGlyHisTyr 100  
QY 451 ATGCAAAACCTTTACCGGAGTCTTATAGGTAGAGTATCCCTCTTTTGTCTCAATG 510  
Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120  
QY 511 CTTCGCGTGAGATTCAACACGAGCATGCAAGAGATTAGAAAGCTATATTAGAGCAATGC 570  
Db 121 LeuGlyValArgPheAsnHisSerMetGlnLysLeuGluSerTyrIleLeuGlnCys 140  
QY 571 TATATCGCTGTGGGCAAAATTTGATGGCGTGTAGCTTAATGGATGTAGTGTGCATT 630  
Db 141 TyrIleAlaValGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
QY 631 ATTGAGGCTTTGATCCTTTAAAGTGGTGAAGTTTGAAGAGCGCTATCAATAAGCT 690  
Db 161 IleGlyGlyPheAspProLeuLysValGlyValLeuLeuGluArgIleAsnLysPro 180  
QY 691 AAAATCGCATGCTTGTATCGCTTTGGCAAGAGGTGGCAGAGCGCCAAAATCAAGA 750  
Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
QY 751 AAATCAAAAGTTGATGCGATTCTTGGTTG 780  
Db 201 LysSerLysValAspAlaIleThrIlePhe 210

RESULT 3

US-10-335-977-6487  
; Sequence 6487, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 6487:
SEQUENCE CHARACTERISTICS:
LENGTH: 210 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...210
SEQUENCE DESCRIPTION: SEQ ID NO: 6487:
US-10-335-977-6487

Alignment Scores:
Pred. No.: 4,59e-107 Length: 210
Score: 1041.00 Matches: 200
Percent Similarity: 97.14% Conservative: 4
Best Local Similarity: 95.24% Mismatches: 6
Query Match: 66.77% Indels: 0
DB: 12 Gaps: 0

US-09-647-661-1 (1-900) x US-10-335-977-6487 (1-210)
QY 151 ATGAAATTTTGGATCAGGAAAAAGAGACAACATTTAAACGAGCGCCATTCTTGCAG 210
DB 1 MetLysPheLeuAspHisGluLysArgGlnLeuLeuAsnGluArgHisSerCysLys 20
QY 211 ATGTTTGACGCCATATGAGTTTCTAGTGAGAAATAGAGAAATCGCTGAATCGCC 270
DB 21 MetPheAspSerHisTyrGluPheSerThrGluLeuGluLeuLeaGluIleAla 40
QY 271 AGGTATCGCAAGCTCTTACACACGACGCCATTTGTGATGGTTACTATAAG 330
DB 41 ArgLeuSerProSerSerTyrAsnThrGlnProTyrHisPheValMetValThrAsnLys 60
QY 331 GATTAAAAAACAATTCAGTCGACAGCTACTTTAATGAAGAAATGATTAAGCGCT 390
DB 61 AspLeuLysLysGlnIleAlaHisSerTyrPheAsnGluLeuMetIleLysSerAla 80
QY 391 TCAGCGTTAATGGTGGATGCTCTTTAAGACCTAGCGAGTTGTACACACGCCATTAC 450
DB 81 SerAlaLeuMetValValCysSerLeuLysProSerGlnLeuLeuProThrSerHisTyr 100
QY 451 ATGCAAAACCTTACCGGAGTCTTATAGGTTAGAGTATCCCTCTTTTGTCAAATG 510
DB 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120

```

CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 6487:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...210  
SEQUENCE DESCRIPTION: SEQ ID NO: 6487:  
US-10-335-977-6487

US-10-335-977-6486  
; Sequence 6486, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 742-4214  
; INFORMATION FOR SEQ ID NO: 6486:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 134 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...134  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6486:  
US-10-335-977-6486

Alignment Scores:  
Pred. No.: 6.47e-62 Length: 134

QY 511 CTGGCGTGAGATTCAACACAGCATGCAAGATTAGAAAGCTATATTTAGAGCAATGC 570  
DB 121 LeuGlyValArgPheAsnHisSerMetGlnLysLeuGluSerTyrIleLeuGluGlnCys 140  
QY 571 TATATCGCTGTGGGGCAAAATTTGTCATGGCGCTGAGCTTAATGGGATTGGATTTGCATT 630  
DB 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
QY 631 ATTGAGGCTTTGATCCCTTTAAAGTGGTCAAGTTTATAGAGAGCGTATCAATAAGCCT 690  
DB 161 IleGlyGlyPheAspProLeuLysValGlyGlnIleLeuGluGluArgIleAsnLysPro 180  
QY 691 AAATCGCATGCTTGCATCGCTTTGGGCAAGAGGTGGCAGAGCGAGCCAAAATCAAGA 750  
DB 181 LysIleValCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerLysLysSerArg 200  
QY 751 AAATCAAAAGTTGATGGGATTACTTGGTTG 780  
DB 201 LysSerLysValAspAlaIleThrIlePhe 210

RESULT 4  
US-10-335-977-6486  
; Sequence 6486, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 742-4214  
; INFORMATION FOR SEQ ID NO: 6486:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 134 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...134  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6486:  
US-10-335-977-6486

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 20, 2004, 19:19:48 ; Search time 67.5 Seconds

(without alignments)

8413.819 Million cell updates/sec

Title: US-09-647-661-1

Perfect score: 1559

Sequence: 1 tgcagattttacagagcgc.....gggggtgttttcagcgcttc 900

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 10.0 , Xgapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO\_epool\_p/US09647661/runat\_20042004.135348.2676/app\_query.fasta\_1.1095  
-DB=SPTREMBL\_25 -QFMT=fastan -SUFFIX=sept -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR SCORE=sept -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HRAPISE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09647661 -CGEN 1.1.86 @runat\_20042004.135348.2676 -NCPU=6 -ICPU=3  
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_25:  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1087	69.7	210	2	O30585 helicobacte

2	1080	69.3	210	2	O8RJ56
3	1077	69.1	210	2	O9FBD5
4	1077	69.1	210	2	O9FBD8
5	1077	69.1	210	2	O8VTN5
6	1076	69.0	210	2	O9FBD0
7	1076	68.9	210	2	O9FBD1
8	1074	68.9	210	2	O9FBE7
9	1074	68.9	210	2	O8RJ83
10	1073	68.8	210	2	O8RJ84
11	1073	68.8	210	2	O9FBC9
12	1073	68.8	210	2	O8VTP5
13	1073	68.8	210	2	O8VTN9
14	1072	68.8	210	2	O9FBD3
15	1072	68.8	210	16	O25608
16	1069	68.6	210	2	O9FBD4
17	1069	68.6	210	2	O8RSS2
18	1068	68.5	210	2	O8VTP7
19	1067	68.4	210	2	O8RJES
20	1067	68.4	210	2	O8VTP1
21	1067	68.4	210	2	O9FBE5
22	1065	68.3	210	2	O8VTP2
23	1065	68.3	210	2	O9FBE6
24	1065	68.3	210	2	O9FBE1
25	1065	68.3	210	2	O8VTN7
26	1062	68.1	210	2	O8VTP6
27	1059	67.9	210	2	O8VTP3
28	1058	67.9	210	2	O9FBE3
29	1058	67.9	210	2	O9FBD2
30	1057	67.8	210	2	O8VTN6
31	1057	67.8	210	2	O9FBE0
32	1057	67.8	210	2	O9FBE4
33	1057	67.8	210	2	O9FBE9
34	1056	67.7	210	2	O9AHD6
35	1056	67.7	210	2	O9AHF8
36	1056	67.7	210	2	O9FBE8
37	1054	67.6	210	2	O8RJ4A
38	1052	67.5	210	2	O9FBD6
39	1051.5	67.4	209	2	O8RJF5
40	1049	67.3	210	2	O9RN84
41	1047	67.2	210	2	O9FBE2
42	1047	67.2	210	2	O9FBE2
43	1041	66.8	210	16	O92KP7
44	1040	66.7	210	2	O9FBD0
45	1040	66.7	210	2	O9RN86

#### ALIGNMENTS

##### RESULT 1

ID	O30585	PRELIMINARY;	PRT;	210 AA.
AC	O30585			
DT	01-JAN-1998	(TRENBLrel. 05, Created)		
DT	01-JAN-1998	(TRENBLrel. 05, Last sequence update)		
DT	01-JUN-2003	(TRENBLrel. 24, Last annotation update)		
DE	NADPH-linked flavin nitroreductase.			
GN	RDXA.			
OS	Helicobacter pylori (Campylobacter pylori).			
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;			
OC	Helicobacteraceae; Helicobacter.			
OX	NCBI_TaxID=210;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=HP500;			
RX	MEDLINE=98283705; PubMed=9622362;			
RA	Goodwin A., Kersault D., Sisson G., Veldhuizen van Zanten S.J.,			
RA	Berg D.E., Hoffman P.S.;			
RT	"Metronidazole resistance in Helicobacter pylori is due to null			
RT	mutations in a gene (rdxa) that encodes an oxygen-insensitive NADPH			
RT	nitroreductase."			
RL	Mol. Microbiol. 28:383-393(1998).			
DR	EMBL; AF012552; AAC46349.1; "			
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.			

DR GO:0006118; P:electron transport; IEA.  
DR InterPro: IPR002086; Aldehyde dehydr.  
DR InterPro: IPR000415; Nitroreductase.  
DR Pfam: PF00881; Nitroreductase; 1.  
DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
SQ SEQUENCE 210 AA; 24123 MW; EF06SC8BA3F0EBBI CRC64;

Alignment Scores:  
Pred. No.: 6,91e-101 Length: 210  
Score: 1087.00 Matches: 210  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 69.72% Indels: 0  
DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x 030585 (1-210)

QY 151 ATGAAATTTTGGATCAGGAAAAAGACAACTATTAAACGAGCGCCATCTTGCAG 210  
DB 1 MetLysPheLeuAspGlnGlyArgGlnLeuLeuAenGluArgHisSerCysLys 20  
QY 211 ATGTTTGACGCAATTAGTTTCTAGTGAAGATTAGAAATCGCTGAATCGCC 270  
DB 21 MetPheAspSerHisTyrGluPheSerSerGluGluLeuGluAlaGluAla 40  
QY 271 AGGCTATCGCAAGCTCTTCAACACGAGCGATGGCATTTTGTGATGTTACTATAAG 330  
DB 41 ArgLeuSerProSerSerTyrAsnThrGlnProPheHisPheValMetValThrAsnLys 60  
QY 331 GATTTAAAAACAAATTCGATGTCACAGCTACTTTATGAAAGAAATGATTAAGCGCT 390  
DB 61 AspLeuLysLysGlnAlaValHisSerTyrPheAenGluGluMetIleLysSerAla 80  
QY 351 TCAGGTTAAATGGTGTATGCTCTTAAAGCTAGCGAGTTGTTACACAGCGCCATTAC 450  
DB 81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100  
QY 451 ATGCAAAACCTTTACCCGAGCTTTATAAGTTAGATGATCCCTTTCTTTCCTCAATG 510  
DB 101 MetGlnAenLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120  
QY 511 CTGGCGTGTGAGATTCAACACGAGCATGCAAGATAGAAAGCTATATTTAGAGCAATGC 570  
DB 121 LeuGlyValArgPheAenHisSerMetGlnArgLeuGluSerTyrIleLeuGlnGlnCys 140  
QY 571 TATATCGCTGTGGGCAAAATTTGCATGGCGTGGATTAATGGGATGATAGTTGCATT 630  
DB 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
QY 631 ATTGAGCGCTTTGATCCTTTAAAGTGGTGAAGTTTATAGAGCGCTATCAATAAGCCT 690  
DB 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluArgIleAsnLysPro 180  
QY 691 AAAATCGCATGCTTGCATGCTTTGGCAAGAGGGTGGCAGAGCGCAAAATCAAGA 750  
DB 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
QY 751 AAATCAAAAGTTGATCGATTACTTGGTTG 780  
DB 201 LysSerLysValAspAlaIleThrTrpLeu 210

RESULT 2  
Q8RJ56 PRELIMINARY; PRT; 210 AA.  
AC Q8RJ56;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE RdxA protein.  
GN RDXA  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
OC Helicobacteraceae; Helicobacter.

OX NCBI\_TaxID=210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=190A, and 190C;  
RA Bereswill S., Krainick C., Herrmann L., Kist M.;  
RT "Absence of rdxA Mutations in High-level Metronidazole Resistant  
RT Clinical Isolates of Helicobacter pylori."  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ305350; CAC83802.1; -;  
DR EMBL: AJ305351; CAC83803.1; -;  
DR GO: 0016491; F:oxidoreductase activity; IEA.  
DR GO: 0006118; P:electron transport; IEA.  
DR InterPro: IPR002086; Aldehyde dehydr.  
DR InterPro: IPR000415; Nitroreductase.  
DR Pfam: PF00881; Nitroreductase; 1.  
DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
SQ SEQUENCE 210 AA; 24067 MW; 8581158BB47A88D CRC64;

Alignment Scores:  
Pred. No.: 3,51e-100 Length: 210  
Score: 1080.00 Matches: 208  
Percent Similarity: 99.52% Conservative: 1  
Best Local Similarity: 99.05% Mismatches: 1  
Query Match: 69.28% Indels: 0  
DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x Q8RJ56 (1-210)

QY 151 ATGAAATTTTGGATCAGGAAAAAGACAACTATTAAACGAGCGCCATCTTGCAG 210  
DB 1 MetLysPheLeuAspGlnGlyArgGlnLeuLeuAenGluArgHisSerCysLys 20  
QY 211 ATGTTTGACGCAATTAGTTTCTAGTGAAGATTAGAAATCGCTGAATCGCC 270  
DB 21 MetPheAspSerHisTyrGluPheSerSerGluGluLeuGluAlaGluAla 40  
QY 271 AGGCTATCGCAAGCTCTTCAACACGAGCGATGGCATTTTGTGATGTTACTATAAG 330  
DB 41 ArgLeuSerProSerSerTyrAsnThrGlnProPheHisPheValMetValThrAsnLys 60  
QY 331 GATTTAAAAACAAATTCGATGTCACAGCTACTTTATGAAAGAAATGATTAAGCGCT 390  
DB 61 AspLeuLysLysGlnIleAlaValHisSerTyrPheAenGluGluMetIleLysSerAla 80  
QY 391 TCAGCGTGTGAGATTCAACACGAGCATGCAAGATAGAAAGCTATATTTAGAGCAATGC 450  
DB 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProHisGlyHisTyr 100  
QY 451 ATGCAAAACCTTTACCCGAGCTTTATAAGTTAGATGATCCCTTTCTTTCCTCAATG 510  
DB 101 MetGlnAenLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120  
QY 511 CTGGCGTGTGAGATTCAACACGAGCATGCAAGATAGAAAGCTATATTTAGAGCAATGC 570  
DB 121 LeuGlyValArgPheAenHisSerMetGlnArgLeuGluSerTyrIleLeuGlnGlnCys 140  
QY 571 TATATCGCTGTGGGCAAAATTTGCATGGCGTGGATTAATGGGATGATAGTTGCATT 630  
DB 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
QY 631 ATTGAGCGCTTTGATCCTTTAAAGTGGTGAAGTTTATAGAGCGCTATCAATAAGCCT 690  
DB 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluArgIleAsnLysPro 180  
QY 691 AAAATCGCATGCTTGCATGCTTTGGCAAGAGGGTGGCAGAGCGCAAAATCAAGA 750  
DB 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
QY 751 AAATCAAAAGTTGATCGATTACTTGGTTG 780  
DB 201 LysSerLysValAspAlaIleThrTrpLeu 210

RESULT 3

ID	Q9FBD5	PRELIMINARY;	PRT;	210 AA.
AC	Q9FBD5;			
DT	01-VAR-2001 (trEMBLrel. 16, Created)			
DT	01-VAR-2001 (trEMBLrel. 16, Last sequence update)			
DT	01-JUN-2003 (trEMBLrel. 24, Last annotation update)			
DE	RdXA.			
GN	RDXA.			
OS	Helicobacter pylori (Campylobacter pylori).			
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;			
OC	Helicobacteraceae; Helicobacter.			
NCBI_TaxID=210;				
RN	[1]_SEQUENCE FROM N.A.			
RP	STRAIN=356A1-2;			
RX	MEDLINE=20356933; PubMed=10898705;			
RA	Solca N.M., Bernasconi M.V., Piffaretti J.C.;			
RT	"Mechanism of metronidazole resistance in Helicobacter pylori:			
RT	comparison of the rdxA gene sequences in 30 strains.";			
RL	Antimicrob. Agents Chemother. 44:2207-2210(2000).			
RN	[2]_SEQUENCE FROM N.A.			
RP	STRAIN=356A1-2;			
RA	Maggi-Solca N., Piffaretti J.C.;			
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBSJ databases.			
DR	EMBL; AF180415; AAC01072.1; "			
DR	GO: GO:0016491; F-oxidoreductase activity; IEA.			
DR	GO: GO:0006118; P-electron transport; IEA.			
DR	InterPro: IPR002086; Aldenhyde dehydr.			
DR	InterPro: IPR000415; Nitroreductase.			
DR	Pfam: PF00881; Nitroreductase; 1.			
DR	PROSITE; PS00070; ALDEHYDE DEHYDR CYG; 1.			
SQ	SEQUENCE 210 AA; 24053 MW; 3B1AF30DF61FAA2 CRC64;			
Alignment Scores:				
Pred. No.:	7,05e-100	Length:	210	
Score:	1077.00	Matches:	207	
Percent Similarity:	99.52%	Conservative:	2	
Best Local Similarity:	98.57%	Mismatches:	1	
Query Match:	69.08%	Indels:	0	
DB:	2	Gaps:	0	
US-09-647-661-1 (1-900) x Q9FBD5 (1-210)				
Qy	151	ATGAAATTTTGGATCAGAAAAAAGAACAACTATTAAACGAGCGCCATTCTTCAAG	210	
Db	1	MetLysPheLeuAspGlnGluLysargGlnLeuLeuAenGluAUGHisSerCysLys	20	
Qy	211	ATGTTTGACAGCAATATGAGTTTCTAGTGAAGATTAGAGAAATCGCTGAATCGCC	270	
Db	21	MetPheAspSerHisTyrGluPheSerThrGluLeuGluGluIleAlaGluIleAla	40	
Qy	271	AGCCTATCGCAAGCTCTTACACACGAGCGCATGGCACTTTGTGATGGTTACTATAAG	330	
Db	41	ArgLeuSerProSerSerTyrAsnThrGlnProTPhHisPheValMetValThrAsnLys	60	
Qy	331	GATTTAAAAACAAATTCGATCGACGACTCTTTAATGAAGAAATGATATAAGCGCT	390	
Db	61	AspLeuLysLysGlnIleAlaValHisSerTyrPheAsnGluGluMetIleLysSerAla	80	
Qy	391	TCAGCGTTAATGTGTGTATGCTTTTAAGACCTACGAGTTGTTACACACGCGCATAC	450	
Db	81	SerAlaLeuMetValValCysSerLeuargProSerGluLeuLeuLeuProHisGlyHisTyr	100	
Qy	451	ATCAAAACCTTTTACCCGAGTCTTATAAGGTTAGAGTGATCCCTCTTTTGTCCAATG	510	
Db	101	MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet	120	
Qy	511	CTTGGCGTGAGATTCACACACGACATGCAAGATTAGAAGCTATATTTAGACCAATGC	570	
Db	121	LeuGlyValArgPheAsnHisSerMetGlnLysLeuGluSerTyrIleLeuGluGlnCys	140	
Qy	571	TATATCGCTGTGGGCAAAATTTGCATGGCGCTGAGCTTAATGGGATTTGGATTTGCATT	630	

Db	141	TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle	160
Qy	631	ATTGAGGCTTTGATCTCTTTAAAAAGTGGGTGAAGTTTGAAGAGCGGTATCAATAAGCCT	690
Db	161	IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro	180
Qy	691	AAAAATCGATCGTTGATGCTTTGGGCAAGAGGGTGCACAGCGCCAAAATCAAGA	750
Db	181	LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg	200
Qy	751	AAATCAAAGTTGATCGGATTACTTGGTTG	780
Db	201	LysSerLysValAspAlaValThrTyrLeu	210
RESULT 4			
Q9FBD8 PRELIMINARY; PRT; 210 AA.			
ID	Q9FBD8	AC	Q9FBD8
DT	01-MAR-2001	(TrEMBLrel. 16, Created)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)	
DE	RdxA	(Oxygen-insensitive NADPH nitroreductase).	
GN	RDxA		
OS	Helicobacter pylori (Campylobacter pylori).		
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;		
OC	Helicobacteraceae; Helicobacter.		
OX	NCBI_TaxID=210;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=338A;		
RX	MEDLINE=20356933; PubMed=10898705;		
RA	Solca N.M., Bernasconi M.V., Piffaretti J.C.;		
RT	"Mechanism of metronidazole resistance in Helicobacter pylori:		
RT	comparison of the rdxA gene sequences in 30 strains.";		
RL	Antimicrob. Agents Chemother. 44:2207-2210(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=338A;		
RA	Maggi-Solca N., Piffaretti J.-C.;		
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=313;		
RA	Marais A., Bilardi C., Cantet F., Megraud F.;		
RT	"Characterization of genes involved in the metronidazole resistance in		
RT	nine strains of Helicobacter pylori isolated before and after		
RT	eradication treatment.";		
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF180410; AAG01069.1; -		
DR	EMBL; AF323009; AAL37282.1; -		
DR	EMBL; AF323008; AAL37281.1; -		
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.		
DR	GO; GO:0006118; P:electron transport; IEA.		
DR	InterPro; IPR002096; Aldehyde dehydr.		
DR	InterPro; IPR000415; Nitroreductase.		
DR	Pfam; PF00891; Nitroreductase; 1.		
DR	PROSITE; PS00070; ALDEHYDE DEHYDR CYS; 1.		
SQ	SEQUENCE 210 AA; 24057 MW; 61B1D871026A081 CRC64;		
Alignment Scores:			
Pred. No.:	7,05e-100	Length:	210
Score:	1077.00	Matches:	208
Percent Similarity:	99.05%	Conservative:	0
Best Local Similarity:	99.05%	Mismatches:	2
Query Match:	69.08%	Indels:	0
DB:	2	Gaps:	0
US-09-647-661-1 (1-900) x Q9FBD8 (1-210)			
Qy	151	ATCAAAATTTTGGATCAGAAAAGAGACACTATTAAAGAGCCCATCTTGTCAAG	210
Db	1	MetLysPheLeuAspGlnGlyLysArgGlnLeuLeuAsnGluArgHisSerCysLys	210

QY 211 ATGTTTGACAGCCATTATGAGTTTCTAGTGAAGAAATAGAGAAATCGCTGAAATCGCC 270  
 DB 21 MetPheaspSerHisTyrGluPheSerSerThrGluLeuGluGluLeuAlaGluLeuAla 40  
 QY 271 AGCTATCGCCAGCTCTTACACACAGCAGCCATGCGATTTGTGATGCTTACTAATAAG 330  
 DB 41 ArgLeuSerProSerTyrAsnThrGlnProTyrPheValMetValThrAsnLys 60  
 QY 331 GATTAAAAAACAATAATGCGAGTCACAGCTACTTTAATGAAGAAATGATTAAAGCGCT 390  
 DB 61 AspLeuLysLysGlnLeuAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80  
 QY 391 TCAGCGTAAATGTTGATGCTCTTAAAGACCTAGCAGAGTTGTTACACACGCGCCATTAC 450  
 DB 81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100  
 QY 451 ATGCAAAACCTTACCGGAGTCTTATAAGTTAGAGTATGCTCTTCTTTGCTCAAAATG 510  
 DB 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120  
 QY 511 CTGGCGTGAGATTCAACACAGCATGCAAGAGTATAGAAAGCTATATTTAGAGCAATGC 570  
 DB 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140  
 QY 571 TATATCGCTGCGGCAATTTGCATGGCGTGAGCTTAATGGAGTTGGATGCTGCTTGCATT 630  
 DB 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
 QY 631 ATTGGAGCTTTGATCCTTTAAAGTGGTGAAGTTTGAAGAGCGTATCAATAAGCCT 690  
 DB 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluArgIleAsnLysPro 180  
 QY 691 AAAATCGATCTTGATCGCTTTGGCAAGAGGTGGCAGAGCGGCAAGGCAAAATCAAGA 750  
 DB 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
 QY 751 AAATCAAAAGTTGATGCGATTACTTGGTTG 780  
 DB 201 LysSerLysValAspAlaIleThrIleu 210  
 RESULT 5  
 Q8VTN5 PRELIMINARY; PRT; 210 AA.  
 ID Q8VTN5  
 AC Q8VTN5  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Oxygen-insensitive NADPH nitroreductase.  
 GN RDXA.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=196;  
 RA Marais A., Billard C., Cantet F., Megraud F.;  
 RT "Characterization of genes involved in the metronidazole resistance in  
 RT nine strains of Helicobacter pylori isolated before and after  
 RT eradication treatment."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF323018; AAL37291.1; -;  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR002086; Aldehyde dehydr.  
 DR InterPro; IPR000415; Nitroreductase.  
 DR Pfam; PF00881; Nitroreductase; 1.  
 DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; 1.  
 SQ SEQUENCE 210 AA; 24095 FW; E37E2986090FBB52 CRC64;  
 Alignment Scores: 7.05e-100 Length: 210  
 Pred. No.:

Score: 1077.00 Matches: 207  
 Percent Similarity: 99.52% Conservative: 2  
 Best Local Similarity: 98.57% Mismatches: 1  
 Query Match: 69.08% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-647-661-1 (1-900) x Q8VTN5 (1-210)  
 QY 151 ATGAAATTTTGGATTCAGGAAAAAAGAGACACTATTAAACGAGGCCCTTCTTGCAAG 210  
 DB 1 MetLysPheLeuAspGlnGluLysArgArgGlnLeuAsnGluArgHisSerCysLys 20  
 QY 211 ATGTTTACAGCCATTATGAGTTTCTAGTGAAGAAATAGAGAAATCGCTGAAATCGCC 270  
 DB 21 MetPheaspSerHisTyrGluPheSerSerGluLeuGluGluLeuAlaGluLeuAla 40  
 QY 271 AGCTATCGCCAGCTCTTAAACACACAGCAGCATGCGATTTGTGATGCTTACTAATAAG 330  
 DB 41 ArgLeuSerProSerTyrAsnThrGlnProTyrPheValMetValThrAsnLys 60  
 QY 331 GATTAAAAAACAATAATGCGAGTCACAGCTACTTTAATGAAGAAATGATTAAAGCGCT 390  
 DB 61 AspLeuLysLysGlnLeuAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80  
 QY 391 TCAGCGTAAATGTTGATGCTCTTAAAGACCTAGCAGAGTTGTTACACACGCGCCATTAC 450  
 DB 81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100  
 QY 451 ATGCAAAACCTTACCGGAGTCTTATAAGTTAGAGTATGCTCTTCTTTGCTCAAAATG 510  
 DB 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheSerGlnMet 120  
 QY 511 CTGGCGTGAGATTCAACACAGCATGCAAGAGTATAGAAAGCTATATTTAGAGCAATGC 570  
 DB 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140  
 QY 571 TATATCGCTGCGGCAATTTGCATGGCGTGAGCTTAATGGAGTTGGATGCTGCTTGCATT 630  
 DB 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
 QY 631 ATTGGAGCTTTGATCCTTTAAAGTGGTGAAGTTTGAAGAGCGTATCAATAAGCCT 690  
 DB 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluArgIleAsnLysPro 180  
 QY 691 AAAATCGATCTTGATCGCTTTGGCAAGAGGTGGCAGAGCGGCAAGGCAAAATCAAGA 750  
 DB 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
 QY 751 AAATCAAAAGTTGATGCGATTACTTGGTTG 780  
 DB 201 LysSerLysValAspAlaIleThrIleu 210  
 RESULT 6  
 Q9FBF0 PRELIMINARY; PRT; 210 AA.  
 ID Q9FBF0  
 AC Q9FBF0  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE RDXA.  
 GN RDXA.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=160A;  
 RA MEDLINE=20356933; PubMed=10898705;  
 RA Solca N.M., Bernasconi M.V., Piffaretti J.C.;  
 RT "Mechanism of metronidazole resistance in Helicobacter pylori:  
 RT comparison of the rdxA gene sequences in 30 strains."  
 RL Antimicrob. Agents Chemother. 44:2207-2210(2000).

```

RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=160A;
RA Maggi-Solca N., Piffaretti J.-C.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF180398; AAG01057.1; -.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR000415; Nitroreductase.
DR Pfam; PF00881; Nitroreductase; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
SQ SEQUENCE 210 AA; 24158 MW; 2DB724AFC2F89DBF CRC64;

Alignment Scores:
Pred. No.: 8.89e-100 Length: 210
Score: 1076.00 Matches: 208
Percent Similarity: 99.52% Conservative: 1
Best Local Similarity: 99.05% Mismatches: 1
Query Match: 69.02% Indels: 0
DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x Q9FBD1 (1-210)
QY 151 ATGAAATTTTGGATCAGGAAAAGAGACAACTATTAAACGAGCGCATCTTGCAG 210
Db 1 MetLysPheLeuAspGlnGluLysArgGlnLeuLeuAsnGluArgHisSerCysLys 20
QY 211 ATGTTTGACAGCCATTATGAGTTTCTAGTGAAGAATAGAGAAATCGCTGAATCGCC 270
Db 21 MetPheAspSerHisTyrGluPheSerSerGluLeuGluGluLeuAla 40
QY 271 AGGCTATCGCCAGCTCTTACAAACGAGCGCATGGCATTTTGTGATCGTTACTATAAG 330
Db 41 ArgLeuSerProSerSerTyrAsnThrGlnProTrpHisPheValMetValThrAsnLys 60
QY 331 GATTTAAAAAACAATTCAGTGCACAGCTACTTTAATGAAGAATGATTAAGAGCGCT 390
Db 61 AspLeuLysLysGlnLeuAlaAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80
QY 391 TCAGCGTTAATCGTGTATGCTTTTAAAGTGGTGAAGTTTAAAGAGCGGTATCAATAAGCCT 450
Db 81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100
QY 451 ATGCAAAACCTTTACCGGAGTCTTATAGGTTAGAGTATCCCTTCTTTGCTCAATG 510
Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120
QY 511 CTGGCGTGAGATTCAACCAAGCATGCAAGATTAAGAGCTATATTTAGAGCAATGC 570
Db 121 LeuGlyValArgPheAsnHisSerMetGlnLysLeuLeuSerTyrIleLeuGluGlnCys 140
QY 571 TATATCGCTGTGGGCAATTTGCGAGCTTAATAGGTTAGAGTATCCCTTCTTTGCTCAATG 630
Db 141 TyrIleAlaValGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160
QY 631 ATTGGAGCGCTTTGATCCCTTTAAAGTGGTGAAGTTTAAAGAGCGGTATCAATAAGCCT 690

RESULT 7
Q9FBD1 PRELIMINARY; PRT; 210 AA.
AC Q9FBD1;
DT 01-MAR-2001 (TrEMBLrel. 16, C-created)

```



```

Db 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro 180
QY 691 AAAATCGCATGCTTGATCGCTTTGGCAAGAGGGTGGCAGAGCGCAAAAATCAAGA 750
Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200
QY 751 AAATCAAAAGTTCATCGGATTAAGTGGTGG 780
Db 201 LysSerLysValAspAlaIleThrTrpLeu 210

RESULT 8
Q9FBE7 PRELIMINARY; PRT; 210 AA.
AC Q9FBE7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE RDXA.
GN RDXA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
CX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22A;
RX MEDLINE=20356933; PubMed=10898705;
RA Solca N.M., Bernasconi M.V., Piffaretti J.C.;
RT "Mechanism of metronidazole resistance in Helicobacter pylori:
RL Antimicrob. Agents Chemother. 44:2207-2210(2000)."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=22A;
RA Maggì-Solca N., Piffaretti J.-C.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF180401; AG01060.1; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR000415; Nitroreductase.
DR Pfam; PF00881; Nitroreductase; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR. CYS; 1.
SQ SEQUENCE 210 AA; 24097 MW; 6B5BFF71CD4F7EC CRC64;

Alignment Scores:
Pred. No.: 1.42e-99 Length: 210
Score: 1074.00 Matches: 207
Percent Similarity: 99.05% Conservative: 1
Best Local Similarity: 98.57% Mismatches: 2
Query Match: 68.89% Indels: 0
DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x Q9FBE7 (1-210)

QY 151 ATGAATTTTGGATCAGGAAAAAGAACAACTATTAAACGAGCGCATCTTCTCAAG 210
Db 1 MetLysPheLeuAspGlnGluLysArgGlnLeuLeuAsnGluHisSerCysLys 20
QY 211 ATGTTGACAGCCATTAGAGTTTCTAGTGAAGAAATGAGAAATCGCTGAATCGCC 270
Db 21 MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluIleAlaGluIleAla 40
QY 271 AGGCTATCGCAAGCTCTTCAACACAGCGCATGCGATTTCGTGATGGTTACTATAAG 330
Db 41 ArgLeuSerProSerSerTyrAsnThrGlnProThrPheValMetValThrAsnLys 60
QY 331 GATTAAAAACAAATTCGAGTCACAGCTACTTTAATGAAGAAATGATTAACACGCT 390
Db 61 AspLeuLysLysGlnIleAlaValHisSerTyrPheAsnGluGluMetIleLysSerAla 80
QY 391 TCAGCGTTAATGGTGGTATGCTCTTTAAGACCTAGCGAGTTGTATACCACGCGCCATTAC 450

```

```

Db 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProHisGlyHisTyr 100
QY 451 ATGCAAAACCTTTACCGGAGTCTTTAAGGTTAGAGTGATCCCTCTCTTTTGTCTCAATG 510
Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheThrGlnMet 120
QY 511 CTTGGCGTGAGATTCAACACACAGCATGCAAGATTAGAAGCTATATTTAGAGCAATGC 570
Db 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140
QY 571 TATATCGCTCTGGGCAAAATTTGCATGGCGGTGAGCTTAATGGATGGATAGTTGCATT 630
Db 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160
QY 631 ATTGGAGGCTTTCATCCTTTAAAGTGGGTGAAGTTTGAAGAGCGGTATCAATAAGCCT 690
Db 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro 180
QY 691 AAAATCGCATGCTTGATCGCTTTGGCAAGAGGGTGGCAGAGCGCAAAAATCAAGA 750
Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200
QY 751 AAATCAAAAGTTCATCGGATTAAGTGGTGG 780
Db 201 LysSerLysValAspAlaIleThrTrpLeu 210

RESULT 9
Q8RJ83 PRELIMINARY; PRT; 210 AA.
AC Q8RJ83
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE RDXA protein.
GN RDXA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
CX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=345A; and 345C;
RA Bereswill S., Krainick C., Herrmann L., Kist M.;
RT "Absence of rdxA Mutations in High-level Metronidazole Resistant
RL Clinical Isolates of Helicobacter pylori."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ305352; CAC83804.1; -.
DR EMBL; AJ305353; CAC83805.1; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR000415; Nitroreductase.
DR Pfam; PF00881; Nitroreductase; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR. CYS; 1.
SQ SEQUENCE 210 AA; 24039 MW; 6F8104871026B981 CRC64;

Alignment Scores:
Pred. No.: 1.42e-99 Length: 210
Score: 1074.00 Matches: 207
Percent Similarity: 99.05% Conservative: 1
Best Local Similarity: 98.57% Mismatches: 2
Query Match: 68.89% Indels: 0
DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x Q8RJ83 (1-210)

QY 151 ATGAATTTTGGATCAGGAAAAAGAACAACTATTAAACGAGCGCATCTTCTCAAG 210
Db 1 MetLysPheLeuAspGlnGluLysArgGlnLeuLeuAsnGluHisSerCysLys 20
QY 211 ATGTTGACAGCCATTAGAGTTTCTAGTGAAGAAATGAGAAATCGCTGAATCGCC 270

```

Db 21 MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluGluLeuAlaGluIleAla 40  
 QY 271 AGGCTATCGCCAGCTCTTACACACGACGCCATGGCATTTGTGTAGTGTACTATAAG 330  
 Db 41 ArgLeuSerProSerTyrAsnThrGlnProTrpHisPheValMetValThrAsnLys 60  
 QY 331 GATTAAAAAACAATAATGTCAGTGCACAGCTACTTTAATGAAGAAATGATTAAGCGCT 390  
 Db 61 AspLeuLysLysGlnIleAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80  
 QY 391 TCAGCGTAAATGCTGATCTCTTAAGACCTAGCAGAGTTGTACACACGCCCATAC 450  
 Db 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProHisGlyHisTyr 100  
 QY 451 ATGCAAACTTACCGGAGTCTTATAAGTTAGAGTATCCCTCTTCTTGTCTCAAAG 510  
 Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120  
 QY 511 CTGGCGTGAATTCACACACGATGCAAGATTAGAAAGCTATATTTAGAGCAATGC 570  
 Db 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140  
 QY 571 TATATCGCTGGGGCAATTCGATGGCGGTGAGCTTAATGGAGTGTGATAGTTCATT 630  
 Db 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
 QY 631 ATTGGAGCTTTCATCCCTTTAAAGTGGTGAAGTTTGAAGAGCGTATCAATAAGCCT 690  
 Db 161 IleGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro 180  
 QY 691 AAAATCGCATCTGATCGCTTTGGGCAAGAGGTGCGCAGACGCGCAAAATCAAGA 750  
 Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
 QY 751 AAATCAAAAGTTCGATCTACTTGGTTG 780  
 Db 201 LysSerLysValAspAlaIleThrTripleu 210

## RESULT 10

Q8RJCA ID Q8RJCA PRELIMINARY; PRT; 210 AA.  
 AC Q8RJCA;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE RdxA protein.  
 DE RDXA.  
 GN RDXA.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=053A, and 053C;  
 RA Bereswill S., Krainick C., Herrmann L., Kist M.;  
 RT "Absence of rdxA Mutations in High-Level Metronidazole Resistant  
 RT Clinical Isolates of Helicobacter pylori."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ305348; CA83800.1;  
 DR EMBL; AJ305349; CA83801.1;  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR002086; Aldehyde dehydr.  
 DR InterPro; IPR000415; Nitroreductase.  
 DR Pfam; PF00881; Nitroreductase; 1  
 DR PROSITE; PS00070; ALDEHYDE\_DEHYD CYS; 1.  
 SQ SEQUENCE 210 AA; 24151 MW; 4BD0113D746F131A CRC64;

## Alignment Scores:

Pred. No.: 1.79e-99 Length: 210  
 Score: 1073.00 Matches: 207  
 Percent Similarity: 99.05% Conservative: 1  
 Best Local Similarity: 98.57% Mismatches: 2

Query Match: 68.83% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-647-661-1 (1-900) x Q8RJCA (1-210)

QY 151 ATGAAATTTTGGATCAGGAAAAAAGAACAACTATTAAACGAGCGCATTTCTTGAAG 210  
 Db 1 MetLysPheLeuAspGlnGluLysArgGlnLeuLeuAsnGluArgHisSerCysLys 20  
 QY 211 ATGTTTGACGCCATTATAGTCTTTCTAGTGAAGAAATAGAGAAATCGCTGAATCGCC 270  
 Db 21 MetPheAspSerHisTyrGluPheSerSerGluLeuGluGluIleAlaGluIleAla 40  
 QY 271 AGGCTATCGCAAGCTCTTACACACGACGATCGCATTTTGTGATGGTGTACTATAAG 330  
 Db 41 ArgLeuSerProSerSerTyrAsnThrGlnProTrpHisPheValMetValThrAsnLys 60  
 QY 331 GATTAAAAAACAATAATGTCAGTGCACAGCTACTTTAATGAAGAAATGATTAAGCGCT 390  
 Db 61 AspLeuLysLysGlnIleAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80  
 QY 391 TCAGCGTAAATGCTGATCTCTTAAGACCTAGCAGAGTTGTACACACGCCCATAC 450  
 Db 81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProTyrGlyHisTyr 100  
 QY 451 ATGCAAACTTACCGGAGTCTTATAAGTTAGAGTATCCCTCTTCTTGTCTCAAATG 510  
 Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheThrGlnMet 120  
 QY 511 CTGGCGTGAATTCACACACGATGCAAGATTAGAAAGCTATATTTAGAGCAATGC 570  
 Db 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140  
 QY 571 TATATCGCTGGGGCAATTCGATGGCGGTGAGCTTAATGGAGTGTGATAGTTCATT 630  
 Db 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
 QY 631 ATTGGAGCTTTCATCCCTTTAAAGTGGTGAAGTTTGAAGAGCGTATCAATAAGCCT 690  
 Db 161 IleGlyLysPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro 180  
 QY 691 AAAATCGCATCTGATCGCTTTGGGCAAGAGGTGCGCAGACGCGCAAAATCAAGA 750  
 Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
 QY 751 AAATCAAAAGTTCGATCTACTTGGTTG 780  
 Db 201 LysSerLysValAspAlaIleThrTripleu 210

RESULT 11  
 Q8RJCA ID Q8RJCA PRELIMINARY; PRT; 210 AA.  
 AC Q8RJCA;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE RdxA.  
 DE RDXA.  
 GN RDXA.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=951;  
 RA MEDLINE=20356933; PubMed=10898705;  
 RX Solca N.M., Bernasconi M.V., Fiffaretti J.C.;  
 RT "Mechanism of metronidazole resistance in Helicobacter pylori:  
 RT comparison of the rdxA gene sequences in 30 strains."  
 RL Antimicrob. Agents Chemother. 44:2207-2210(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=951;

RA Maggi-Solca N., Piffaretti J.-C.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF180423; AAC01078.1; -;  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR002086; Aldehyde dehydr.  
 DR InterPro; IPR000415; Nitroreductase.  
 DR Pfam; PF00881; Nitroreductase; I.  
 DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; 1.  
 SQ SEQUENCE 210 AA; 24163 MW; 46C79907237A0207 CRC64;

Alignment Scores:  
 Pred. No.: 1.79e-99 Length: 210  
 Score: 1073.00 Matches: 207  
 Percent Similarity: 99.05% Conservative: 1  
 Best Local Similarity: 98.57% Mismatches: 2  
 Query Match: 68.83% Indels: 0  
 DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x Q9FBC9 (1-210)

```

QY 151 ATGAAATTTTGGATCAGGAAAAAGAGACAACTATTAAACGAGCGCATCTTCGAG 210
Db 1 MetLysPheLeuAspGlnGluLysArgGlnLeuLeuAsnGluArgHisSerCysLys 20
QY 211 ATGTTTCACAGCCATTATGATTTCTAGTCAAGAATTAGAGAAATCGCTGAATCGCC 270
Db 21 MetPheAspSerHisTyrGluPheSerSerGluGluLeuGluGluLeuAla 40
QY 271 AGCTATCGCAGCTCTTCAACACGAGCCATGCGATTTGTGATGCTTACTATAAG 330
Db 41 ArgLeuSerProSerTyrAsnThrGlnProTrpHisPheValIleValThrAsnLys 60
QY 331 GATTAAAAAACAATTCGAGTCACAGCTACTTTAATGAAGAATGATTAAGAGCGCT 390
Db 61 AspLeuLysLysGlnIleAlaValHisSerTyrPheAsnGluGluMetIleLysSerAla 80
QY 391 TCAGCGTTAATGTTGATGCTCTTTAAGACCTTAGAGCTTGTACACGCGCATTTAC 450
Db 81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisSerHisTyr 100
QY 451 ATGCAAAACCTTTACCGGAGTCTTAAAGTTAGAGTATGCTTCTTTGCTCAATG 510
Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120
QY 511 CTGCGGTGAGTCAACACGAGCATGCAAGATTAGAAAGCTATATTTAGAGCAATGC 570
Db 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGlnCys 140
QY 571 TATATCGTGTGGGCAATTTGATCGGCGTGTAGCTTAAAGAGCGGTATGATGCTTGCATT 630
Db 141 TyrIleAlaValGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160
QY 631 ATTGAGGCTTTGATCCTTTAAAAGTGGGTGAAGTTTTAGAGAGCGGTATCAATAAGCCT 690
Db 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluArgIleAsnLysPro 180
QY 691 AAAATCGCATCTTGATCGCTTTGGCGAAGGGTGGCAGAGCGGCAAAAATCAAGA 750
Db 181 LysIleValCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200
QY 751 AAATCAAAAGTTGATGCTTACTTGGTTG 780
Db 201 LysSerLysValAspAlaIleThrTripleu 210

```

RESULT 12

Q8VTP5 PRELIMINARY; PRT; 210 AA.  
 AC Q8VTP5;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Oxygen-insensitive NADPH nitroreductase.

GN RDXA.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=220;  
 RA Marais A., Bilardi C., Cantet F., Megraud P.;  
 RT "Characterization of genes involved in the metronidazole resistance in  
 RT nine strains of Helicobacter pylori isolated before and after  
 RT eradication treatment";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF323006; AAL37279.1; -;  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR002086; Aldehyde dehydr.  
 DR InterPro; IPR000415; Nitroreductase.  
 DR Pfam; PF00881; Nitroreductase; 1.  
 DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; 1.  
 SQ SEQUENCE 210 AA; 24035 MW; D22D60620BC2657 CRC64;

Alignment Scores:  
 Pred. No.: 1.79e-99 Length: 210  
 Score: 1073.00 Matches: 206  
 Percent Similarity: 99.52% Conservative: 3  
 Best Local Similarity: 98.10% Mismatches: 1  
 Query Match: 68.83% Indels: 0  
 DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x Q8VTP5 (1-210)

```

QY 151 ATGAAATTTTGGATCAGGAAAAAGAGACAACTATTAAACGAGCGCATCTTCGAG 210
Db 1 MetLysPheLeuAspGlnGluLysArgGlnLeuLeuAsnGluArgHisSerCysLys 20
QY 211 ATGTTTCACAGCCATTATGATTTCTAGTCAAGAATTAGAGAAATCGCTGAATCGCC 270
Db 21 MetPheAspSerHisTyrGluPheSerSerGluGluLeuGluGluLeuAla 40
QY 271 AGCTATCGCAGCTCTTCAACACGAGCCATGCGATTTGTGATGCTTACTATAAG 330
Db 41 ArgLeuSerProSerTyrAsnThrGlnProTrpHisPheValIleValThrAsnLys 60
QY 331 GATTAAAAAACAATTCGAGTCACAGCTACTTTAATGAAGAATGATTAAGAGCGCT 390
Db 61 AspValLysLysGlnIleAlaValHisSerTyrPheAsnGluGluMetIleLysSerAla 80
QY 391 TCAGCGTTAATGTTGATGCTCTTTAAGACCTTAGAGCTTGTACACGCGCATTTAC 450
Db 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProHisGlyHisTyr 100
QY 451 ATGCAAAACCTTTACCGGAGTCTTAAAGTTAGAGTATGCTTCTTTGCTCAATG 510
Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120
QY 511 CTGCGGTGAGTCAACACGAGCATGCAAGATTAGAAAGCTATATTTAGAGCAATGC 570
Db 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGlnCys 140
QY 571 TATATCGTGTGGGCAATTTGATCGGCGTGTAGCTTAAAGAGCGGTATGATGCTTGCATT 630
Db 141 TyrIleAlaValGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160
QY 631 ATTGAGGCTTTGATCCTTTAAAAGTGGGTGAAGTTTTAGAGAGCGGTATCAATAAGCCT 690
Db 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro 180
QY 691 AAAATCGCATCTTGATCGCTTTGGCGAAGGGTGGCAGAGCGGCAAAAATCAAGA 750
Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200
QY 751 AAATCAAAAGTTGATGCTTACTTGGTTG 780

```

```

Db      201 LysSerLysValAspAlaIleThrTrpLeu 210
|||||
RESULT 13
Q8VTN9
ID      Q8VTN9          PRELIMINARY;      PRT;      210 AA.
AC      Q8VTN9;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      Oxygen-insensitive NADPH nitroreductase.
GN      RDXA.
OS      Helicobacter pylori (Campylobacter pylori).
OC      Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC      Helicobacteraceae; Helicobacter.
OX      NCBI_TaxID=210;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CAS;
RA      Marais A., Bilardi C., Cantet F., Megraud F.;
RT      "Characterization of genes involved in the metronidazole resistance in
RT      nine strains of Helicobacter pylori isolated before and after
RT      eradication treatment.";
RL      Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF232014; AAL37287.1; -.
DR      GO; GO:0016491; F:oxidoreductase activity; IEA.
DR      GO; GO:0006118; P:electron transport; IEA.
DR      InterPro; IPR002086; Aldehyde dehydr.
DR      InterPro; IPR000415; Nitroreductase.
DR      Pfam; PF00881; Nitroreductase; 1.
DR      PROSITE; PS00070; ALDEHYDE DEHYDR. CYS; 1.
SQ      SEQUENCE 210 AA; 24097 MW; 0ADC66F70463B4D0 CRC64;

Alignment Scores:
Pred. No.:      1.79e-99      Length:      210
Score:          1073.00      Matches:      207
Percent Similarity: 98.57%      Conservative: 0
Best Local Similarity: 98.57%      Mismatches: 3
Query Match:    68.93%      Indels:      0
DE:             2           Gaps:          0

US-09-647-661-1 (1-900) x Q8VTN9 (1-210)

QY      151 ATGAAATTTTGGATCAGGAAAAAGAACACACTATTAAACGAGCGCCATTCTTGCAG 210
Db      1 MetLysPheLeuAspGlnGluYsArgGlnLeuLeuAsnGluArgHisSerCysLys 20
|||||
QY      211 ATGTTTGACGCCATTATGAGTTTCTAGTCAAGAATTAGAAGAAATCGCTGAATCGCC 270
Db      21 MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluGluIleAla 40
|||||
QY      271 AGGCTATCGCGAAGCTCTTACACACGCGCATGTCATTTGTCGTTACTATAAG 330
Db      41 ArgLeuSerProSerSerTyrAsnThrGlnProThrHisPheValMetValThrAsnLys 60
|||||
QY      331 GATTTAAAAAACAATTCAGTCCAGCTACTTTAATGAAGAAATGATTAAGCGCT 390
Db      61 AspLeuLysGlnIleAlaAlaHisSerTyrPheAsnGluMetIleLysSerAla 80
|||||
QY      391 TCACGGTTAATGGTATGCTCTTTAAGACCTAGCGATGTTTACACAGCGCCATTAC 450
Db      81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100
|||||
QY      451 ATGCAAAACCTTACCCGAGCTCTATAAGGTAGATGATCCCTTCTTTGCTCAAATG 510
Db      101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheThrGlnMet 120
|||||
QY      511 CTGGCGTGAGATTCAACACAGCATGCAAGATAGAAAGCTATATTTTAGAGCAATGC 570
Db      121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140
|||||
QY      571 TATATCGCTGGGGCAAAATTTGATGGCGGTGAGCTTAATGGGATTCGATTCGATT 630
|||||

Db      141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160
631 ATTCGAGCGCTTGTGATCTTTAAAGTGGTGAAGCTTTTAGAGAGCGCTATCAATAAGCGCT 690
Db      161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro 180
691 AAAATCGCATGCTTTCATCGCTTTGGCGAAGAGGTGGCAGAGCGAGCCAAAATCAAGA 750
Db      181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200
751 AAATCAAAAGTTGATCGCATTCATTCTGTTG 780
Db      201 LysSerLysValAspAlaIleThrTrpLeu 210

RESULT 14
Q9FBD3
ID      Q9FBD3          PRELIMINARY;      PRT;      210 AA.
AC      Q9FBD3;
DT      01-MAR-2001 (TrEMBLrel. 16, Created)
DT      01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      RDXA.
GN      RDXA.
OS      Helicobacter pylori (Campylobacter pylori).
OC      Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC      Helicobacteraceae; Helicobacter.
OX      NCBI_TaxID=210;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=56A;
RA      Solca N.M., Bernasconi M.V., Piffaretti J.C.;
RT      "Mechanism of metronidazole resistance in Helicobacter pylori:
RT      comparison of the rdxA gene sequences in 30 strains.";
RL      Antimicrob. Agents Chemother. 44:2207-2210(2000).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=56A;
RA      Maggi-Solca N., Piffaretti J.-C.;
RL      Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF180417; AAG01074.1; -.
DR      GO; GO:0016491; F:oxidoreductase activity; IEA.
DR      GO; GO:0006118; P:electron transport; IEA.
DR      InterPro; IPR002086; Aldehyde dehydr.
DR      InterPro; IPR000415; Nitroreductase.
DR      Pfam; PF00881; Nitroreductase; 1.
DR      PROSITE; PS00070; ALDEHYDE DEHYDR. CYS; 1.
SQ      SEQUENCE 210 AA; 24110 MW; C77FBEE7B2C36F64 CRC64;

Alignment Scores:
Pred. No.:      2.25e-99      Length:      210
Score:          1072.00      Matches:      205
Percent Similarity: 99.05%      Conservative: 3
Best Local Similarity: 97.62%      Mismatches: 2
Query Match:    68.76%      Indels:      0
DE:             2           Gaps:          0

US-09-647-661-1 (1-900) x Q9FBD3 (1-210)

QY      151 ATGAAATTTTGGATCAGGAAAAAGAACACACTATTAAACGAGCGCCATTCTTGCAG 210
Db      1 MetLysPheLeuAspGlnGluYsArgGlnLeuLeuAsnGluArgHisSerCysLys 20
|||||
QY      211 ATGTTTGACGCCATTATGAGTTTCTAGTCAAGAATTAGAAGAAATCGCTGAATCGCC 270
Db      21 MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluGluIleAla 40
|||||
QY      271 AGGCTATCGCGAAGCTCTTACACACGCGCATGTCATTTGTCGTTACTATAAG 330
Db      41 ArgLeuSerProSerSerTyrAsnThrGlnProThrHisPheValMetValThrAsnLys 60
|||||
QY      331 GATTTAAAAAACAATTCAGTCCAGCTACTTTAATGAAGAAATGATTAAGCGCT 390
Db      61 AspLeuLysGlnIleAlaAlaHisSerTyrPheAsnGluMetIleLysSerAla 80
|||||
QY      391 TCACGGTTAATGGTATGCTCTTTAAGACCTAGCGATGTTTACACAGCGCCATTAC 450
Db      81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100
|||||
QY      451 ATGCAAAACCTTACCCGAGCTCTATAAGGTAGATGATCCCTTCTTTGCTCAAATG 510
Db      101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheThrGlnMet 120
|||||
QY      511 CTGGCGTGAGATTCAACACAGCATGCAAGATAGAAAGCTATATTTTAGAGCAATGC 570
Db      121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140
|||||
QY      571 TATATCGCTGGGGCAAAATTTGATGGCGGTGAGCTTAATGGGATTCGATTCGATT 630
|||||

```

```

Db 61 AspLeuLysGlnIleAlaAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80
QY 391 TCAGCGTAATGGGTATGCTCTTTAAAGACCTTAGCGAGTTGTACACACGCCCAATAC 450
Db 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProHisGlyHisTyr 100
QY 451 ATGCAAAACCTTTACCGGAGTCTTAAGCTTAGAGTGATCCCTCTTTTGCTCAAAATG 510
Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheThrGlnMet 120
QY 511 CTTCGCTGAGATTCAACACAGCATGCAAGATAGAAAGCTATATTTTAGAGCAATGC 570
Db 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140
QY 571 TATATCGTGTGGGCGAAATTTGCATGGCGGTGAGCTTAATGGGATTGGATAGTTGCATT 630
Db 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160
QY 631 ATTGGAGCTTTGATCCCTTTAAAGTGGGTGAAGTTTTAGAAGAGCTATCAATAAGCCT 690
Db 161 IleGlyGlyPheAspProLeuLysValGlyGluIleLeuGluGlnArgIleAsnLysPro 180
QY 691 AAAATCCCATCTTGATCGCTTTGGCAAGAGGGTGGCAGAGCGGCCCAAAATCAAGA 750
Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200
QY 751 AAAATCAAAAGTTGATCGCATTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 780
Db 201 LysSerLysValAspAlaIleThrTrpLeu 210

RESULT 15
O25608 PRELIMINARY; PRT; 210 AA.
AC O25608;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Oxygen-insensitive NAD(P)H nitroreductase.
GN HP0954.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Felschmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547 (1997).
DR EMBL; AE00604; RAD07997.1; -.
DR FIR; B64639; B64639.
DR TIGR; HP0954; -.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR000415; Nitroreductase.
DR Pfam; PF00881; Nitroreductase; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 210 AA; 24068 MW; DFF15D2D102C002C CRC64;

```

Alignment Scores: 2.25e-99 Length: 210  
Pred. No.:

```

Score: 1072.00 Matches: 207
Percent Similarity: 99.05% Conservative: 1
Best Local Similarity: 98.57% Mismatches: 2
Query Match: 68.76% Indels: 0
DB: 16 Gaps: 0

US-09-647-661-1 (1-900) x O25608 (1-210)
QY 151 ATGAATTTTGGATCAGGAAAAAGAACAACTATTAAACGAGCGCATTTCTTCAAG 210
Db 1 MetLysPheLeuAspGlnGluLysArgArgGlnLeuLeuAsnGluArgHisSerCysLys 20
QY 211 ATGTTTGACACCCATTATGAGTTTCTAGTGAAGAAATAGAGAAATCGCTGAATCGCC 270
Db 21 MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluIleAlaGluIleAla 40
QY 271 AGCTATCGCCCAAGCTCTTACAAACGACGACCATGGCATTTTGTGATCGTTACTAATAAG 330
Db 41 ArgLeuSerProSerSerTyrAsnThrGlnProThrPheValMetValThrAspLys 60
QY 331 GATTTAAAAAACAATTCGAGTGCACAGCTACTTTAATGAAGAAATGATTAAGCGCT 390
Db 61 AspLeuLysLysGlnIleAlaAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80
QY 391 TCAGCGTTAATCGTGTATGCTCTTTAAGACTAGCGAGTTGTACACACGCCCATTCAC 450
Db 81 SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100
QY 451 ATGCAAAACCTTTACCGGAGTCTTTAAGGTTAGAGTGATCCCTCTTTTGTCTCAATG 510
Db 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120
QY 511 CTTCGCTGAGATTCAACACAGCATGCAAGATAGAAAGCTATATTTTAGAGCAATGC 570
Db 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140
QY 571 TATATCGTGTGGGCGAAATTTGCATGGCGGTGAGCTTAATGGGATTGGATAGTTGCATT 630
Db 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160
QY 631 ATGAGAGGCTTTGATCCCTTTAAAGTGGGTGAAGTTTTAGAAGAGCTATCAATAAGCCT 690
Db 161 IleGlyGlyPheAspProLeuLysValGlyGluIleLeuGluGlnArgIleAsnLysPro 180
QY 691 AAAATCGCATGCTTGATCGCTTTGGCAAGAGGGTGGCAGAGCGGCCCAAAATCAAGA 750
Db 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200
QY 751 AAAATCAAAAGTTGATCGCATTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 780
Db 201 LysSerLysValAspAlaIleThrTrpLeu 210

```

Search completed: April 20, 2004, 19:28:00  
Job time : 72.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 20, 2004, 19:20:42 ; Search time 24.5 Seconds  
(without alignments)  
7067.133 Million cell updates/sec

Title: US-09-647-661-1  
Perfect score: 1559  
Sequence: 1 tgcagattttacagagagc.....gggggtttttcgaagcgtttc 900

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Rgapop 6.0, Rgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_n2p.model -DEV=xl2  
-C=/cgn2/1/USPTO\_spool\_p/US09647661/runat\_20042004\_135348\_2706/app\_query.fasta\_1.1095  
-DB=PIR\_78 -QFWT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=spt -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09647661@cgn 1 1 38 @runat\_20042004\_135348\_2706 -NCPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREAS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR 78:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1072	68.8	210	2 B64639	oxygen-insensitive
2	1041	66.8	210	2 A71876	probable aldehyde
3	246	15.8	284	1 C64639	prolipo protein dia
4	229	14.7	283	2 B71876	probable prolipo
5	226	14.5	223	2 D90056	hypothetical prote
6	223.5	14.3	221	2 B69809	NAD(P)H-flavin oxi
7	223.5	14.3	221	2 B61949	probable NAD(P)H-f
8	218.5	14.0	220	1 B64114	probable NAD(P)H-f
9	208.5	13.4	221	2 C81155	NAD(P)H nitroreduc
10	203	13.0	222	2 E83700	NAD(P)H-flavin oxi
11	199.5	12.8	217	2 C82436	oxygen-insensitive
12	198.5	12.7	201	2 E81309	nitroreductase Cj1
13	198.5	12.7	217	1 A38686	nitroreductase (EC
14	196.5	12.6	217	1 S08397	nitroreductase (EC

#### ALIGNMENTS

##### RESULT 1

B64639  
oxygen-insensitive NAD(P)H nitroreductase - Helicobacter pylori (strain 26695)  
C;Species: Helicobacter pylori  
C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 11-Jun-1999  
C;Accession: B64639  
R;Comb, J.P.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen, son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujiki, C.; Bowman, C.; Watthey, i Nature 388, 539-547, 1997  
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A;Reference number: A64520, MUID:97394467; PMID:9252185  
A;Accession: B64639  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-210 <TOM>  
A;Cross-references: GB:AE000604; GB:AE000511; NID:g2314086; PIDN:AAD07997.1; PID:g231414  
C;Superfamily: nitroreductase

Alignment Scores:  
Pred. No.: 2.3e-93 Length: 210  
Score: 1072.00 Matches: 207  
Percent Similarity: 99.05% Conservative: 1  
Best Local Similarity: 98.57% Mismatches: 2  
Query Match: 68.76% Indels: 0  
DB: Gaps: 0

US-09-647-661-1 (1-900) x B64639 (1-210)

Qy 151 ATGAATTTTGGATCAGGAAAGAGACAACTATTAAAGAGCGCCATTCTTCCAAG 210  
1 MetLysPheLeuAspGlnGluLysAArgGlnLeuLeuAsnGluAHisSerCysLys 20  
Qy 211 ATCTTTGACAGCCATTATGAGTTTCTAGTGAAGATTAGAGAAATCGCTGAATCGCC 270  
Db 21 MetPheAspSerHisrYrGluPheSerThrGluLeuGluCylleAlaGluIleAla 40

15 194.5 12.5 217 2 AD0573 oxygen-insensitive  
16 191.5 12.3 217 2 H90705 oxygen-insensitive  
17 191.5 12.3 217 2 C85556 oxygen-insensitive  
18 190 12.2 217 2 H71912 probable oxidoredu  
19 188.5 12.1 217 1 I67695 nitroreductase (EC  
20 185.5 11.9 202 2 H69902 nitroreductase hom  
21 182 11.7 217 1 B64600 NAD(P)H-flavin oxi  
22 181.5 11.6 201 2 S75047 drgA protein - Syn  
23 172.5 11.1 209 2 C69783 NADH dehydrogenase  
24 169.5 10.9 321 2 B72384 bacterioferritin c  
25 167.5 10.7 271 2 B98179 probable NAD(P)H n  
26 167 10.7 210 2 B87429 nitroreductase fam  
27 165.5 10.6 213 2 D83929 NAD(P)H nitroreduc  
28 164.5 10.6 209 2 A83108 nitroreductase (im  
29 160 10.3 201 2 B35072 nitroreductase fam  
30 159 10.2 210 2 A81087 NADH oxidase homol  
31 157.5 10.1 201 2 B97940 nitroreductase (EC  
32 153.5 9.8 179 2 G97074 nitroreductase fam  
33 153.5 9.8 356 2 H86808 transcription regu  
34 153 9.8 174 1 C69533 NAD(P)H-flavin oxi  
35 153 9.8 215 2 A75331 nitroreductase - D  
36 145.5 9.3 212 2 C84132 NADH dehydrogenase  
37 144.5 9.3 203 2 D86647 oxidoreductase ybi  
38 144 9.2 210 2 A81451 NADH oxidase homol  
39 143 9.2 213 2 A84250 NADH oxidase limpo  
40 140 9.0 208 2 G97004 nitroreductase fam  
41 139.5 8.9 200 2 E82996 probable nitroredu  
42 137.5 8.8 206 2 D69781 conserved hypothet  
43 129.5 8.3 194 2 C89266 NAD(P)H-flavin oxi  
44 126 8.1 191 1 B99027 NADPH-oxidoreducta  
45 126 8.1 249 2 A82039 hypothetical prote

QY 271 AGGCTATCGCAAGCTCTTCAACACGACGCGCATTTTGATGGTTACTAATAAG 330  
 Db |||||  
 QY 41 ArgLeuSerProSerSerTyAsnThrGlnProTriPhisPheValMetValThrAspLys 60  
 Db |||||  
 QY 331 GATTATAAAACAAATGCGTCCACAGCTACTTTAATGAAGAATGATTAAGCGCT 390  
 Db |||||  
 QY 61 AspLeuLysGlnIleAlaAlaHisSerTyPheAsnGluGluMetIleLysSerAla 80  
 QY 391 TCAGCGTAAATGGTGGTATGCTCTTAAAGACCTAGCGAGTGTGTACACACGCGCCATTAC 450  
 Db |||||  
 QY 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProHisGlyHisTy 100  
 QY 451 ATGCAAAACCTTTACCGGAGCTTTAAGTTAGTAGTATGCCCTCTTTTGTCTCAATG 510  
 Db |||||  
 QY 101 MetGlnAsnLeuTyProGluSerTyLysValArgValIleProSerPheAlaGlnMet 120  
 QY 511 CTTCGGCTGAGATTCAACACAGCATGCAAGATGATAAGAGCTATATTTTAGAGCAATGC 570  
 Db |||||  
 QY 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyIleLeuGluGlnCys 140  
 QY 571 TATATCGCTGGGCGCAAAATTGTCATGGCGGTGAGCTTAATGGGATTGGATAGTGCATT 630  
 Db |||||  
 QY 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
 QY 631 ATTGGAGCTTTGATCCTTTAAAGTGGTGAGCTTTTAGAGAGCGTATCAATAGCCT 690  
 Db |||||  
 QY 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluArgIleAsnLysPro 180  
 QY 691 AAAATCGCATCTTGATCGCTTTGGCAAGAGGTGGCAGAACGCGCAAAATCAAGA 750  
 Db |||||  
 QY 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
 QY 751 AAATCAAACTGTGCGATTACTTGGTTG 780  
 Db |||||  
 QY 201 LysSerLysValAspAlaIleThrTrpLeu 210  
 Db |||||

RESULT 2  
 A:1876  
 C:Probable aldehyde dehydrogenase - Helicobacter pylori (strain J99)  
 A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 11-Jun-1999  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
 A:Reference number: A71800; PMID:99120557; PMID:9923682  
 A:Accession: A71876  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-210 <ARN>  
 A:Cross-references: GB:AE001518; GB:AE001439; NID:G4155454; PIDN:AA006472.1; PID:G415546  
 A:Experimental source: strain J99  
 C:Genetics:  
 C:Gene: jnp0888  
 C:Superfamily: nitroreductase

Alignment Scores:  
 Pred. No.: 1,98e-90 Length: 210  
 Score: 1041.00 MatChes: 200  
 Percent Similarity: 97.14% Conservatives: 4  
 Best Local Similarity: 95.24% Mismatches: 6  
 Query Match: 66.77% Indels: 0  
 DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x A71876 (1-210)

QY 151 ATGAATTTTGGATCAGAAAGACACAACTATTAAACGCGCCATTCTTGAAG 210  
 Db |||||  
 QY 1 MetLysPheLeuAspHisGlnLysArgGlnLeuLeuAsnGluArgHisSerCysLys 20  
 QY 211 ATGTTTGACGCCATTATGAGTTTCTAGTGAAGAATTAGAAGAAATCGCTGAAATCGCC 270

Db 21 MetPheAspSerHisTyGluPheSerSerThrGluLeuGluGluIleAlaGluIleAla 40  
 QY 271 AGGCTATCGCAAGCTCTTCAACACGACGCGCATTTTGATGGTTACTAATAAG 330  
 Db |||||  
 QY 41 ArgLeuSerProSerSerTyAsnThrGlnProTriPhisPheValMetValThrAsnLys 60  
 QY 331 GATTATAAAACAAATGCGTCCACAGCTACTTTAATGAAGAATGATTAAGCGCT 390  
 Db |||||  
 QY 61 AspLeuLysGlnIleAlaAlaHisSerTyPheAsnGluGluMetIleLysSerAla 80  
 QY 391 TCAGCGTAAATGGTGGTATGCTCTTAAAGACCTAGCGAGTGTGTACACACGCGCCATTAC 450  
 Db |||||  
 QY 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProHisGlyHisTy 100  
 QY 451 ATGCAAAACCTTTACCGGAGCTTTAAGTTAGTAGTATGCCCTCTTTTGTCTCAATG 510  
 Db |||||  
 QY 101 MetGlnAsnLeuTyProGluSerTyLysValArgValIleProSerPheAlaGlnMet 120  
 QY 511 CTTCGGCTGAGATTCAACACAGCATGCAAGATGATAAGAGCTATATTTTAGAGCAATGC 570  
 Db |||||  
 QY 121 LeuGlyValArgPheAsnHisSerMetGlnLysLeuGluSerTyIleLeuGluGlnCys 140  
 QY 571 TATATCGCTGGGCGCAAAATTGTCATGGCGGTGAGCTTAATGGGATTGGATAGTGCATT 630  
 Db |||||  
 QY 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
 QY 631 ATTGGAGCTTTGATCCTTTAAAGTGGTGAGCTTTTAGAGAGCGTATCAATAGCCT 690  
 Db |||||  
 QY 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluArgIleAsnLysPro 180  
 QY 691 AAAATCGCATCTTGATCGCTTTGGCAAGAGGTGGCAGAACGCGCAAAATCAAGA 750  
 Db |||||  
 QY 181 LysIleValCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerLysLysSerArg 200  
 QY 751 AAATCAAACTGTGCGATTACTTGGTTG 780  
 Db |||||  
 QY 201 LysSerLysValAspAlaIleThrTrpLeu 210  
 Db |||||

RESULT 3  
 C:64639  
 C:Protein diacylglycerol transferase (EC 2.4.99.-) - Helicobacter pylori (strain :  
 C:Species: Helicobacter pylori  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: C64639  
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.I.  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, I.  
 Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A:Reference number: A64520; PMID:97394467; PMID:9252185  
 A:Accession: C64639  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-284 <TOM>  
 A:Cross-references: GB:AE000604; GB:AE000511; NID:G2314086; PIDN:AA07998.1; PID:G2314  
 C:Superfamily: prolipoprotein diacylglycerol transferase  
 C:Keywords: glycosyltransferase

Alignment Scores:  
 Pred. No.: 3.7e-15 Length: 284  
 Score: 246.00 Matches: 49  
 Percent Similarity: 100.00% Conservatives: 1  
 Best Local Similarity: 98.00% Mismatches: 0  
 Query Match: 15.78% Indels: 0  
 DB: 1 Gaps: 0

US-09-647-661-1 (1-900) x C64639 (1-284)

QY 2 GCGAATTTTACAGAGCGCAGATAGCCAAATCGGGGTTATTTTAAATTTGACATG 61  
 Db |||||  
 QY 235 AlaGluPheTyArgGluProAspSerGlnMetGlyValTyPheLeuAsnLysSerMet 254

QY 62 GGGCAGATTTTAACTTATTTAGTGTGTTTCAATAGGAGTTTATTTATGTTATGCTACA 121  
 Db 255 GlyGlnIleLeuSerLeuPheMetValIleValSerLeuGlyIleLeuLeuTyAlaThr 274  
 QY 122 AAAAATCTTAAAAATAAAGAAATCAA 151  
 Db 275 LysAsnSerLysLysIleLysGluAsnGln 284

RESULT 4  
 B71876  
 probable prolipoprotein diacylglycerol transferase - Helicobacter pylori (strain J99)  
 C:Species: Helicobacter pylori  
 A:Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 18-Jun-1999  
 C:Accession: B71876  
 R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, B.D.; Doig, P.C.; Smith, D.R.;  
 ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
 A:Reference number: A71800; MUID:99120557; PMID:9923692  
 A:Accession: B71876  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-283 <ARN>  
 A:Cross-references: GB:AE001439; NID:G4155454; PIDN:AAD06473.1; PID:G415547  
 A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: lgt  
 C:Superfamily: prolipoprotein diacylglycerol transferase

Alignment Scores:  
 Pred. No.: 1.5e-13 Length: 283  
 Score: 229.00 Matches: 45  
 Percent Similarity: 100.00% Conservatives: 3  
 Best Local Similarity: 93.75% Mismatches: 0  
 Query Match: 14.69% Indels: 0  
 DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x B71876 (1-283)  
 QY 2 GCAGATTTTACAGAGCCAGATAGCAATAGGGGTTTATTTTAAATTTGAGCATG 61  
 Db 235 AlaGluPheTyArgGluProAspSerGlnLeuGlyValTyPheLeuAsnLeuSerMet 254  
 QY 62 GGGCAGATTTTAACTTATTTAGTGTGTTTCAATAGGAGTTTATTTATGTTATGCTACA 121  
 Db 255 GlyGlnIleLeuSerValPheMetValIleValSerLeuGlyIleLeuLeuTyAlaThr 274  
 QY 122 AAAAATCTTAAAAATAAAGAA 145  
 Db 275 LysAsnSerLysLysIleLysGlu 282

RESULT 5  
 D90056  
 hypothetical protein SA2311 [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 01-Mar-2002  
 C:Accession: D90056  
 R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A99758; MUID:21311952; PMID:11418146  
 A:Accession: D90056  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-223 <KUR>  
 A:Cross-references: GB:BA000018; PID:G13702473; PIDN:BAB43614.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA2311

C:Superfamily: nitroreductase

Alignment Scores:  
 Pred. No.: 2.84e-13 Length: 223  
 Score: 226.00 Matches: 55  
 Percent Similarity: 51.21% Conservatives: 51  
 Best Local Similarity: 26.57% Mismatches: 87  
 Query Match: 14.50% Indels: 14  
 DB: 2 Gaps: 5

US-09-647-661-1 (1-900) x D90056 (1-223)  
 QY 196 GCGCATTTCTGCAGATGTTTTCAGACGCCATTTATGATGTTTCTAGTGAAGATTTAGAAGAA 255  
 Db 15 ArgHisAlaThrLysGlnPheAspProGlnLysLysValSerLysGluAspPheGluThr 34  
 QY 256 ATCGCTGAATTCGCCAGCTATCGCAAGCTCTTACACACGCCAGCCATGCGATTGTG 315  
 Db 35 IleLeuGluSerGlyArgLeuSerProSerLeuGlyLeuGluProTrpLysPheVal 54  
 QY 316 ATGGTTTACTAATAGGATTTAAAAACAAATTCGAGTCGACAGCTACTTTAATGAAGAA 375  
 Db 55 ValIleGlnAspGlnAlaLeuArgAspGluLeuLysAlaHisSerTrpGlyAlaAlaLys 74  
 QY 376 ATGATTAAGCGCTTCAGCTTAATGCTGTATGCTCTTTAAGACCTAGCGAGTTGTTA 435  
 Db 75 GlnLeuAspThrAlaSerHisPheValLeuIlePheAlaArgLys-----AsnValThr 92  
 QY 436 CCACACGCCCATTCATCATGCAAAACCTTTACCGGAG-----TCTTATAAGGTTAGAGTG 489  
 Db 93 SerArgSerProTyValGlnHisMetLeuArgAspIleLysLysTyArgGluAlaGlnThr 112  
 QY 490 ATCCCTTCTTTTGTCTCAAAATGCTTGGCGTG-----AGATTCAACACACAGC 534  
 Db 113 IleProAlaValGluGlnLysPheAspAlaPheGlnAlaAspPheHisIleSerAspAsn 132  
 QY 535 ATGCAAGATTTAGAAAGCTATATTTCAGCAATGCTATATCGTGTGCGGCAATTTGC 594  
 Db 133 AspGlnAlaLeuTyAspTrpSerSerLysGlnThrTyIleAlaLeuGlyAsnMetMet 152  
 QY 585 ATGGCGGTGAGCTTAATCGGATTCGATAGTTGCATATTATGGAGGCTTTTCATCTTTAAA 654  
 Db 153 ThrThrAlaAlaLeuLeuGlyIleAspSerCysProMetGluGlyPheSerLeuAspThr 172  
 QY 655 GTGGGTGAAGTTTGAAGAGCGT-----ATCAATAAGCCTAAATCGCATGC 702  
 Db 173 ValThrAspIleLeuAlaAsnLysGlyIleLeuAspThrGluGlnPheGlyLeuSerVal 192  
 QY 703 TTGATCGCTTTGGGCAAGAGGTGGCAGAA---GCGAGCCAAAATCAAGAAAATCAAAA 759  
 Db 193 MetValAlaPheGlyTyArgGlnGlnAspProProLysAsnLysThrArgGlnAlaTy 212  
 QY 760 GTTGATCGCATTCATTCTGGTTG 780  
 Db 213 GluAspValIleGluTrpVal 219

RESULT 6  
 B69809  
 NAD(P)H-flavin oxidoreductase homolog yfko - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: B69809  
 R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bert  
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; C  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabbet, C.; Ferrari, E  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
 Y., M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlo  
 A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser



akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.P.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: B69809  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-221 <KON>  
A:Cross-references: GB:Z39108; GB:AL009126; NID:g2633055; PIDN:CAB12612.1; PID:g2633107  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yfkO  
C:Superfamily: nitroreductase

Alignment Scores:  
Pred. No.: 4,89e-13 Length: 221  
Score: 223.50 Matches: 57  
Percent Similarity: 49.09% Conservative: 51  
Best Local Similarity: 25.91% Mismatches: 91  
Query Match: 14.34% Indels: 21  
DB: 2 Gaps: 4  
US-09-647-661-1 (1-900) x B69809 (1-221)

QY 175 AGAAGACAACTATTA-----AACGAGCGCCATCTTCAAGATGTTTGACAGCCAT 225  
Db 5 LysThrGlnIleLeuAspAlaTyrAsnPhelgHisIalThrLysGluPheAspProAsn 24  
QY 226 TATGAGTTTCTAGTGAAGATTAGAGAAATCGCTGAATCGCCAGGCTATCGCCAAAGC 285  
Db 25 LysLysValSerAspSerAspGluPheIleLeuGluThrGlyArgLeuSerProSer 44  
QY 286 TCTTACACACGCGCCATGTCATTTGTGTGTTCTACTAATAGGATTAAAGAAACAA 345  
Db 45 SerLeuGlyLeuGluProTyrPheValValGlnAsnProGluPheArgGluLys 64  
QY 346 ATTCAGTGCACAGCTACTTAAATCAAGAAATGATTAAAGCGCTTCAGCGTTAATGGTG 405  
Db 65 LeuArgGluTyrThrTrpGlyAlaGlnLysGlnLeuProThrAlaSerHisPheValLeu 84  
QY 406 GTATGCTCTTAAGACCTAGCGAGTTGTATACACAGCCCATACATGCAAAACCTTTAC 465  
Db 85 IleLeuAlaArgThrAlaLysAspIleLysTyrAsnAlaAspTyrIleLysArgHisLeu 104  
QY 466 CCGAGTCTTAAAGGTTAGATGATCCCTCTTTGCTCAATGCTTGGCGTGAGATTC 525  
Db 105 LysGlu-----ValLysGlnMetProGlnAspValTyrGluGlyTyrLeuSerLys 121  
QY 526 AACACACATGCAAGATTAGAAGCTATATTTAGAG----- 564  
Db 122 ThrGluGlyPheGlnLysAsnAspLeuHisLeuGluSerAspArgThrLeuPheAsp 141  
QY 565 -----CAATGCTATATCGCTGGGCAAAATTTGCATGGCGGTGAGCTTAATG 612  
Db 142 TrpAlaSerLysGlnThrTyrIleAlaLeuGlyAsnMetMetThrAlaAlaGlnIle 161  
QY 613 CGATTGGATGATTGAGGCTTTGATCCTTTAAAGTGGGTGAAGTTTAAAG 672  
Db 162 GlyValAspSerCysProIleGluGlyPheGlnTyrAspHisIleHisArgIleLeuGlu 181  
QY 673 GAG-----CGTATCAATAAGCCTAAATCGCATCTGATCGCTTTGGGCAAG 720  
Db 182 GluGluGlyLeuLeuGluAsnGlySerPheAspIleSerValMetAlaAlaPheGlyTyr 201  
QY 721 AGGTTGGCAGAGCGCAAAATCAAGAAATCAAGAAATGATTGATCTGTTG 780  
Db 202 ArgValArgAspProArgProLysThrArgSerAlaValGluAspValValLysTrpVal 221  
RESULT 7  
E81949  
probable NAD(P)H-flavin oxidoreductase NMA1015 [imported] - *Neisseria meningitidis* (stra  
C:Species: *Neisseria meningitidis*

C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: E81949  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More  
; Holtroyd, S.; Jags, K.; Leather, S.; Mungall, K.; Quail, M.A.; Rajandran  
Natus 404, 502-505, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: E81949  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-221 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84284.1; PID:g73797  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA1015  
C:Superfamily: nitroreductase

Alignment Scores:  
Pred. No.: 4,89e-13 Length: 221  
Score: 223.50 Matches: 58  
Percent Similarity: 47.79% Conservative: 50  
Best Local Similarity: 25.66% Mismatches: 95  
Query Match: 14.34% Indels: 23  
DB: 2 Gaps: 4  
US-09-647-661-1 (1-900) x E81949 (1-221)

QY 151 ATGAAATTTTGGATCAGGAAAGAAAGACAACTATTAAGCGCCATCTTGCAG 210  
Db 1 MetThrValLeuAspArgGlnGlnValLeuSerAlaPheLysAsnArgLysSerCysArg 20  
QY 211 ATGTTTGACAGCCATATGAGTTTCTAGTGAAGAAATAGAGAAATCGCTGAATCGCC 270  
Db 21 HisTyrAspAlaAlaArgLysIleSerAlaGluAspPheGlnPheIleLeuGluLeuGly 40  
QY 271 AGCTATCGCCAGCTCTTACACAGCGCCATGCGATTTGTCATGGTACTTAATAG 330  
Db 41 ArgLeuSerProSerValGlySerGluProTyrGlnPheValValGlnAsnPro 60  
QY 331 GATTTAAAGAAACAAATTCGATGCGACAGCTATTAATGAAGAAATGATTAAAGCGCT 390  
Db 61 GluIleArgGlnAlaIleLysProPheSerTrpGlyMetAlaAspAlaLeuAspThrAla 80  
QY 391 TCAGCGTTAATGGTGGTATGCTCTTTAAGACCTAGCGAGTTGTACACAGCGCCATTAC 450  
Db 81 SerHisLeuValValPheLeuAlaLysLys-----AsnAlaArgPhe 94  
QY 451 ATGCAAAACCTTTACCGGAGTCTTATAAGGTTAGAGTGATC-----CCTTCTTT 501  
Db 95 AspSerProPheMetLeuGluSerLeuLysArgAlaGlyValThrGluProAspAlaMet 114  
QY 502 GTCATAATGCTTGGC-----GTGAGATTCAACCAACAGC 534  
Db 115 AlaLysSerLeuAlaArgTyrGlnAlaPheGlnAlaAspAspIleLysIleLeuAspAsp 134  
QY 535 ATGCAAGATTAGAAAGCTATATTTAGACCAATCTATCGCTGGGCAAAATTTGC 594  
Db 135 SerArgAlaLeuPheAspTrpCysArgGlnThrTyrIleAlaLeuGlyAsnMetMet 154  
QY 595 ATGGCGGTGAGCTTAATGGATGATGATGTCATTTATGGAGGCTTTCATCTTTAA 654  
Db 155 ThrGlyAlaAlaAlaGlyIleAspSerCysProValGluGlyPheAsnTyrAlaAsp 174  
QY 655 GTGGGTGAAGTTTGAAGAGCGTATCAAT-----AAGCCTAAATTCGA 699  
Db 175 MetGluArgValLeuSerGlyGlnPheGlyLeuPheAspAlaAlaGluTrpGlyValSer 194  
QY 700 TGGTTGATCGCTTGGCAAGAGGCTGGCAGAGCGAGCCAAAATCAAGAAAATCAAAA 759  
Db 195 ValAlaAlaThrPheGlyTyrArgValGlnGluIleAlaThrLysAlaArgProLeu 214  
QY 760 GTTGATCGCATCTACTGG 777

Db 215 GluGluThrValIleTrp 220  
RESULT 8  
B64114  
probable NAD(P)H-flavin oxidoreductase (EC 1.6.6.-) - Haemophilus influenzae (strain Rd  
C;Species: Haemophilus influenzae  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: B64114  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A.;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630; PMID:7542800  
A;Accession: B64114  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-220 <IGR>  
A;Cross-references: GB:U32807; GB:L42023; NID:91574725; PIDN:AAC22926.1; PID:g1574733; T  
A;Note: named as homolog to a protein from Vibrio fischeri  
C;Superfamily: nitroreductase  
C;Keywords: NAD; oxidoreductase  
Alignment Scores:  
Pred. No.: 1.45e-12 Length: 220  
Score: 218.50 Matches: 59  
Percent Similarity: 48.89% Conservative: 51  
Best Local Similarity: 26.22% Mismatches: 94  
Query Match: 14.02% Indels: 21  
DB: 1 Gaps: 4  
US-09-647-661-1 (1-900) x B64114 (1-220)  
QY 151 ATGAAATTTTGGATCAGGAAAAAGACACAACTATTAAACGAGCGCCATCTTGGCAAG 210  
Db 1 MetThrGlnLeuThrArgGluGlnValLeuLeuPheHisGlnArgSerThrArg 20  
QY 211 ATGTTTGACAGCATTATGATGTTTCTAGTGAAGATTAGACAAATCCCTGAATCGCC 270  
Db 21 TyrTyrAspProThrLysLysIleSerAspGluAspPheGluCysIleLeuGluCysGly 40  
QY 271 AGGCTATCGCAAGCTCTTACACAGCAGCCATGATTTGTGATGTTACTAATAAG 330  
Db 41 ArgLeuSerProSerValGlySerGluProTyrLysPheLeuValIleGlnAsnLys 60  
QY 331 GATTAAAAAACAATTCGAGTCACAGCTACTTAAATGAAGAAATGATTAAGCGCT 390  
Db 61 ThrLeuArgGluLysMetLysProPheSerTrpGlyWetIleAsnGlnLeuAspAsnCys 80  
QY 391 TCAGCGTAAATGTTGGTATGCTCTTAAAGACCTAGCGAGTGTGTACACAGCGCCATTAC 450  
Db 81 SerHisLeuValIleLeuAlaLysLys-----AsnAlaArgTyr 94  
QY 451 ATGCAAAACCTTTACCGGAGTCTTAAAGTTAGATGATCCCTCTTTTGTCAAATG 510  
Db 95 AspSerProPhePheValAspValMetAlaArgLysGlyLeuAsnAlaGluGlnGlnGln 114  
QY 511 CTGGCGTGAGATCAACACAGCATGCAAGATTAGAAGCTATATTTTAGN----- 564  
Db 115 AlaAlaLeuThrLysTyrLysAlaLeuGlnGluAspMetLysLeuLeuGluAsnAsp 134  
QY 565 -----CAATGCTATATCGCTGTGGGCAAAATTTGCATG 597  
Db 135 ArgThrLeuPheAspTrpCysSerLysGlnThrTyrIleAlaLeuAlaAsnMetLeuThr 154  
QY 598 GCGGTGAGCTTAATGGATTGGATGATGTCATTATTGGAGGCTTTGATCCTTTAAAGTG 657  
Db 155 GlyAlaSerAlaLeuGlyIleAspSerCysProIleGluGlyPheHisTyrAspLysMet 174  
QY 658 GGTGAAGCTTTA--GAAGAGCGGTATCAATAAGCTAAA-----ATCGCATGCTTG 705  
Db 175 AsnGluCysLeuAlaGluGluGlyLeuPheAspProGlnGluTyrAlaValSerValAla 194

QY 706 ATCGCTTTGGCAAGAGCGTGGCAGAGCGAGCGCAAAAATCAAGAAATCAAAAGTTGAT 765  
Db 195 AlaThrPheGlyTyrArgSerArgAspIleAlaLysLysSerArgLysGlyLeuAspGlu 214  
QY 766 GCGATTACTTGGTTG 780  
Db 215 ValValLysTrpVal 219  
RESULT 9  
C81155  
NAD(P)H nitroreductase, probable NMB0804 [imported] - Neisseria meningitidis (strain MC  
C;Species: Neisseria meningitidis  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C;Accession: C81155  
R;Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizzia, M.  
Science 287, 1809-1815, 2000  
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V  
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A;Reference number: A81000; MUID:20175755; PMID:10710307  
A;Accession: C81155  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-221 <TET>  
A;Cross-references: GB:AF002435; GB:AF002098; NID:g7226049; PIDN:AAF41217.1; PID:g72260  
A;Experimental source: serogroup B, strain MC58  
C;Genetics:  
A;Gene: NMB0804  
C;Superfamily: nitroreductase  
Alignment Scores:  
Pred. No.: 1.29e-11 Length: 221  
Score: 208.50 Matches: 58  
Percent Similarity: 46.46% Conservative: 47  
Best Local Similarity: 25.66% Mismatches: 98  
Query Match: 13.37% Indels: 23  
DB: 2 Gaps: 5  
US-09-647-661-1 (1-900) x C81155 (1-221)  
QY 151 ATGAAATTTTGGATCAGGAAAAAGACAACTATTAAACGAGCGCCATCTTGGCAAG 210  
Db 1 MetThrValLeuSerLysGluGlnValLeuSerAlaPheLysAsnArgLysSerCysArg 20  
QY 211 ATGTTTGACAGCCATTATGAGTTTCTAGTGAAGAAATGAGAAATCGCTGAATCGCC 270  
Db 21 HisTyrAspAlaAlaArgLysIleSerAlaGluAspPheGlnPheIleLeuGluLeuGly 40  
QY 271 AGGCTATCGCAAGCTCTTACACAGCAGCCATGCGCATTTTGTGATGTTTACTAATAAG 330  
Db 41 ArgLeuSerProSerValGlySerGluProTyrLysPheIleValIleValGlnAsnPro 60  
QY 331 GATTAAAAAACAATTCGAGTCACAGCTACTTAAATGAAGAAATGATTAAGCGCT 390  
Db 61 GluIleArgGlnAlaIleLysProPheSerTrpGlyMetAlaAspAlaLeuAspThrAla 80  
QY 391 TCAGCGTAAATGTTGGTATGCTCTTAAAGACCTAGCGAGTGTGTACACAGCGCCATTAC 450  
Db 81 SerHisLeuValValPhe-----LeuAlaLysLysAsnAlaArgSerAspSerProPhe 98  
QY 451 ATGCAAAACCTTTACCGGAGTCTTATAAGGTTTAGA-----GTG 489  
Db 99 MetLeu-----GluSerLeuLysArgGlyValThrGluProAspAlaVal 114  
QY 490 ATCCCTTTCTTTGCT-----CAATGCTGGCGTGAGATTCACACACAGC 534  
Db 115 AlaLysSerLeuAlaArgTyrGlnAlaPheGlnAlaAspPheIleLysIleLeuAspAsp 134  
QY 535 ATGCAAGATTAGAAAGCTATATTTAGACCAATGCTATATCGCTGTGGGCAAAATTCG 594  
Db 135 SerArgAlaLeuPheAspTrpCysArgGlnThrTyrIleAlaLeuAlaAsnMetMet 154

595 ATCGGGCTGAGCTTAATGGGATTGGATGCGATTGTCATTATTGGAGCGCTTTGATCCTTTAAAA 654  
 155 ThrGlyAlaMetAlaGlyIleAgsPseCysProValGluGlyPheAenTyRAlaGlu 174  
 655 GTGGGTGAAGTTTTAGAGAGCGGTATCAAT-----AAGCCTAAATCGCA 699  
 175 MetGluArgIleLeuSerGlyGlnPheGlyLeuPheAspAlaAlaGluTrpGlyValSer 194  
 700 TGCTTGATCGCTTTGGCGAAGGGTGGCAGAGCGCGCAAAAATCAAGAAAATCAAAA 759  
 195 ValAlaAlaThrPheGlyTyRArgValGlnGluIleAlaThrLysAlaArgProLeu 214  
 760 GTTGATCGGATTACTGG 777  
 215 GluGluThrValIleTrp 220

RESULT 10  
 E83700  
 NAD(P)H-flavin oxidoreductase BH0405 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 01-Mar-2002  
 C:Accession: E83700  
 R:R.Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira-  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: E83700  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-222 <STO>  
 A:Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BA04124.1; GSPDB:GNOC  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH0405  
 C:Superfamily: nitroreductase

Alignment Scores:  
 Pred. No.: 4,27e-11 Length: 222  
 Score: 203.00 Matches: 56  
 Percent Similarity: 50.46% Conservative: 54  
 Best Local Similarity: 25.69% Mismatches: 88  
 Query Match: 13.02% Indels: 20  
 DB: 2 Gaps: 7

US-09-647-661-1 (1-900) x E83700 (1-222)

Qy 175 AGAAGCAACTATTAACGAG-----CGCATCTTCGAGATGTTTGACGCCAT 225  
 Db 8 LysGluGluIleLeuAenAlaTyRArgPheArgHisAlaThrLysGluPheAgsProAsn 27  
 Qy 226 TATGAGTTTTCTAGTGAAGAATAGAGAATCGCTGAAATCGCAGGCTATCGCCAAAGC 285  
 Db 28 LysThrIleProAlaAgsPheGlnPheIleLeuGluThrGlyArgLeuSerProSer 47  
 Qy 286 TCTTACAACAGCGCCATCGCATTTTGTGATGGTTACTAATAAGGATTTAAAAAACAA 345  
 Db 48 SerPheGlyMetGluProTrpArgPheIleValGlnAsnGluLysLeuArgAspLys 67  
 Qy 346 ATTGCAGTGCACACTACTTTAATGAAGAAATGATTAAGACGGCTTCAGCGTTTAATGGTG 405  
 Db 68 IleLysAsnThrSerTrpGlyAlaAlaGlyLysLeuProGluAlaSerHisPheValMet 87  
 Qy 406 GTATGCTCTTAAAGACCTAGGAGTTGTTACCACAGCGCCATTACATGCAAAACCTTTAC 465  
 Db 88 IleLeuAlaThrHisLeuAspThrArgPheAspSerAlaTyRLeuLysGluGlnPhe 107  
 Qy 466 CCGGAGCTTATAGGTTAGAGTATCCCTCTTTTCTCAATGCTTGGCGTG----- 519  
 Db 108 SerAspPheProGluAspArgLeu-----SerGlnTyRLeuGlyMetIleGlu 123  
 Qy 520 AGATTC---AACCACGACGATCAAGATTAGAA-----AGCTATATT 558  
 Db 88 LysGluGluIleLeuAenAlaTyRArgPheArgHisAlaThrLysGluPheAgsProAsn 27

Db 124 GlnPheGlnLysHisAspPheLysLeuLeuAspGlnGluArgPheLeuAsnAspTrpAla 143

Qy 559 TTAGAGCAATGCTATATCGTGTGGGCAAAATTTCATGGCGTGAGCTTAATGGGATTG 618

Db 144 IleArgGlnThrTrpIleProLeuAlaAsnMetMetThrAlaAlaGluIleGlyLe 163

Qy 619 GATAGTTCATTATTGAGGCTTTGATCCCTTTAAAGCTGGTGAAGTTTAA--GAGAG 675

Db 164 AspSerCysProIleGluGlyPheAsnValGluArgMetAsnGluLeuAlaGlu 183

Qy 676 CGTATCAATAAGCCT-----AAAATCGCATGCTTGATCGCTTTGGGCAAGGGGTG 726

Db 184 GlyLeuLeuGluAspGlyHisPheThrIleSerValMetValAlaPheGlyTyrArgLys 203

Qy 727 GCGAGCGAGCCAAATAATCAAGAAATCAAAGTTGATCGATTACTTGGTTG 780

Db 204 LysGluProGlnProLysThrArgArgProTyrAspAspValValLysTrpIle 221

RESULT 11

C82436

oxygen-insensitive NAD(P)H nitroreductase VCA0637 [imported] - Vibrio cholerae (strain C/Species: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C/Accession: C82436

R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.I.; Dodson, R.J.; Chardson, D.; Ercolae, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, L. R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; MUID: 20406833; PMID:10952301

A/Accession: C82436

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-217 <HEI>

A/Cross-references: GB:AE004393; GB:AE003853; NID:c9658042; PIDN:AAF96538.1; GSPDB:GN000

A/Experimental source: serogroup O1; strain N16961; biotype E1 Tor

C/Genetics:

A/Gene: VCA0637

A/Map position: 2

C/Superfamily: nitroreductase

Alignment Scores:

Pred. No.: 9.13e-11 Length: 217

Score: 199.50 Matches: 55

Percent Similarity: 45.36% Conservative: 33

Best Local Similarity: 28.35% Mismatches: 53

Query Match: 12.80% Indels: 53

DB: 2 Gaps: 6

US-09-647-661-1 (1-900) x C82436 (1-217)

Qy 181 CAACTATTAAACGAGCCCACTTTGCAAGATGTTGACGCCATTATGATTTTCTAGT 240

Db 5 GlnAlaSerGlnSerArgTyrSerThrLysAlaPheAspAlaSerArgLysLeuSerGlu 24

Qy 241 GAAGAAATGAGAAATCGCTGAATCGCCAGGCTATCGCCAGCTCTTACACACGAG 300

Db 25 GlnGlnValAlaAspLeuLysGluLeuValArgMetSerAlaSerSerValAsnSerGln 44

Qy 301 CCATGGCATTTTGTGATGGTTTACTATAAGATTTAAAAAACAATTCGATGCACAGC 360

Db 45 ProTrpHisPheIleLeuAlaGlySerAspGluGlyThrArgIleAlaLysAlaThr 64

Qy 361 TAC-----TTTATGAGAAGATGATTAAAGCGCTTCAGCGCTTAATGCTGTA 408

Db 65 GlnGlyGlyPheSerPheAsnGluArgLysIleLeuAspAlaSerHisValMetValPhe 84

Qy 409 TGC-----TCTTTAAGACCTAGCGAGTTGTTACCA----- 438

Db 85 CysAlaLysThrSerIleAspAlaTyrLeuLeuSerLeuLeuAspAsnGluAspLys 104

Qy 439 -----CAGCGC-----CATTTACATG 453

Db 105 AspGlyArgPheAlaAsnGluAlaLysThrGlyMetHisGlyAlaArgSerTyrPhe 124  
 QY 454 CAAACCTTTACCGGAGTCTTAAAGTTAGAGTGATCCCTTCTTTGCTCAAAAGCTT 513  
 Db 125 ValAsnLeuHisArgGlu-----130  
 QY 514 GCGGTGAGATTCAACACGACGATGCAAGATTAGAAAGCTATATTTTAGAGCAATGCTAT 573  
 Db 131 -----AsnLeuAsnAspAlaGluHisTrpMetGlnLysGlnValTyr 144  
 QY 574 ATCGCTGTGGGCAAAATTCATGCGGTGAGCTTAATGGATTTGATGATGCTATATT 633  
 Db 145 LeuAsnValGlyThrLeuLeuGlyAlaAlaMetGlyIleAspAlaValProIle 164  
 QY 634 GGAGGCTTTCATCTTTAAAGTGGGTGAAGTTTGAAGAG 675  
 Db 165 GluGlyPheAsp-----AlaGlnValLeuAsnGlu 174

RESULT 12  
 E81309  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C:Accession: E81309  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrer  
 Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
 A:Reference number: A81250; MUID:20150912; PMID:10688204  
 A:Accession: E81309  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-201 <PAR>  
 A:Cross-references: GB:AL139077; GB:AL111168; NID:G6968444; PIDN:CAB73321.1; PID:G696850  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: Cj1064; Cj1066

Alignment Scores:  
 Pred. No.: 1,13e-10 Length: 201  
 Score: 198.50 Matches: 52  
 Percent Similarity: 50.26% Conservative: 46  
 Best Local Similarity: 26.67% Mismatches: 68  
 Query Match: 12.73% Indels: 29  
 DB: 2 Gaps: 6

US-09-647-661-1 (1-900) x E81309 (1-201)  
 QY 169 GAAAGAGAGACAACTATTAAACGCGCATTTGCAAGATGTTTGACAGCCATTAT 228  
 Db 2 LysLysGluLeuGluIlePheSerThrArgTyrSerCysArgAsnPhelLysAsnGlu--- 20  
 QY 229 GAGTTTCTAGTGAAGATTAGAGAAATCGCTGAATCCCGAGGTATCGCCAAAGCTCT 288  
 Db 21 LysLeuLysLysGluAspLeuAsnSerIleLeuGluIleAlaArgLeuSerProSerSer 40  
 QY 289 TACAACAGCGAGCATGCGATTTTGTGATGTTACTTAATAAGGATTAAACAAAT 348  
 Db 41 LeuGlyLeuGluProIlePheIleValValGlnAspGlnLysArgLysGluGluLeu 60  
 QY 349 GCAGTGACAGCTACTTTAATGAAGAAATG---ATTAAAGCGCTTCAGCGTTTAATGGTG 405  
 Db 61 Ser-----LysIleCysAsnGlnLysHisValLysAspCysAlaAlaLeuIleIle 78  
 QY 406 GTATGCTCT-----TTAGACCTACGAGTTG 432  
 Db 79 IleIleSerArgLeuAspPheLeuAspTyrPheGluGlnLysLeuArgLysArgAspMet 98  
 QY 433 TTACCACAGCGCCATTACATGCAAAACCTTTACCGGAGTCTTATAGGTTAGAGTGATC 492  
 Db 99 -----SerGluThrGluMetGlnLysArgLeu 107  
 QY 493 CCTTCTTTGCTCAAAATGCTTGGCGTGAGATTCAACACGACGATGCAAGATTAGAAAGC 552

Db 108 AspThrTyrMetProPheLeuLys-----SerLeuAsnGlnGluGlnLysIleSer 124  
 QY 553 TATATTTTAGAGCAATGCTATATCGCTGTGGGGCAAAATTTGCATGGCGTGAGCTTAATG 612  
 Db 125 TyrAlaArgGluGlnAlaHisIleAlaLeuAlaSerIleLeuTyrSerAlaAsnAlaLeu 144  
 QY 613 GGATTGGTACTTCGATTTATTTGGAGCTTTGATCCTTTAAAGTGGGTGAAGTTTGAAG 672  
 Db 145 AsnIleAlaSerCysThrIleGlyPheAspLysGlnLysLeuAspSerTyrLeuSer 164  
 QY 673 GAGCGTATCAATAACCTCAAAATCGCATGCTTGATCGCTTTGGGC 717  
 Db 165 LeuAspIleGlnLysGluArgSerSerLeuValValAlaLeuGly 179

RESULT 13  
 A38686  
 C:Species: Enterobacter cloacae  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A38686  
 R:Bryant, C.; Hubbard, L.; McElroy, W.D.  
 J. Biol. Chem. 266, 4126-4130, 1991  
 A:Title: Cloning, nucleotide sequence, and expression of the nitroreductase gene from I  
 A:Reference number: A38686; MUID:91154203; PMID:1999406  
 A:Accession: A38686  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-217 <BRY>  
 A:Cross-references: GB:M63808; GB:M37085; NID:G148361; PIDN:AAA62801.1; PID:G148362  
 C:Superfamily: nitroreductase  
 C:Keywords: oxidoreductase

Alignment Scores:  
 Pred. No.: 1,14e-10 Length: 217  
 Score: 198.50 Matches: 51  
 Percent Similarity: 48.09% Conservative: 37  
 Best Local Similarity: 27.87% Mismatches: 56  
 Query Match: 12.73% Indels: 39  
 DB: 1 Gaps: 5

US-09-647-661-1 (1-900) x A38686 (1-217)  
 QY 193 GAGCGCCATTCTGCAAGATGTTGACAGCCATTATGAGTTTCTAGTGAAGATTAGAA 252  
 Db 9 LysArgHisSerThrLysAlaPheAspAlaSerLysLysLeuThrAlaGluAlaGlu 28  
 QY 253 GAAATCGCTGAAATCGCAGGCTATCGCCAAAGCTTTTACAAACAGCGAGCCATGCTTTT 312  
 Db 29 LysIleLysThrLeuLeuGlnTyr-SerProSerThrAsnSerGlnProTrpHisPhe 48  
 QY 313 GTGATGGTTACTAATAAGGATTAAACAAACAAATTCGAGTGCACAGC----- 360  
 Db 49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyThrTyr 68  
 QY 361 TACTTTAATGAAGAAATGATTAAAGCGCTTTCAGCGTTTAATGGTGGTATGCTCTTTAGA 420  
 Db 69 ValPheAsnGluArgLysMetLeuAspAlaSerHisValValPheCysAla----- 86  
 QY 421 CCTAGCGAGTTGTACACAGCGCCATTACATGCAAAACCTTTACCCGGAGTCTTATAG 480  
 Db 87 -----LysThrAlaMetAspAspAlaTrpLeu 95  
 QY 481 GTTAGAGTGATC-----CCTTCTTTT 501  
 Db 96 GluArgValValAspGlnGluAlaAspGlyArgPheAsnThrProGluAlaLysAla 115  
 QY 502 GCTCAAAATGCTTGGCGTGAGATTCAACACAGCATGCAAGATT----- 546  
 Db 116 AlaAsnHisLysGlyArgThrTyrPheAlaAspMetHisArgValAspLeuLysAsp 135  
 QY 547 GAAAGCTATATTTAGACCAATGCTATATCCTCTGCGGGCAAAATTTGTCATGGCGTGAGC 606

Db 136 AspGlnTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeuGlyValGly 155  
 QY 607 TTAATGGATTGGATTCATTCATTTATGGAGGCTTTCATCCITTAAGTGGGTGAAGTT 666  
 Db 156 AlaMetGlyLeuAspAlaValProIleGluGlyPheAsp-----AlaAlaIle 171  
 QY 667 TTAGAAGAG 675  
 Db 172 LeuAspGlu 174

## RESULT 14

S08397  
 nitroreductase (EC 1.6.6.-) - Salmonella typhimurium  
 C:Species: Salmonella typhimurium  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S08397  
 R:Watanabe, M.; Ishidate Jr., M.; Nohmi, T.  
 Nucleic Acids Res. 18, 1059, 1990  
 A:Title: Nucleotide sequence of Salmonella typhimurium nitroreductase gene.  
 A:Reference number: S08397; MUID:90192100; PMID:2179862  
 A:Molecule type: DNA  
 A:Accession: S08397  
 A:Residues: 1-217 <WAT>  
 A:Cross-references: EMBL:X17250; NID:g47792; PIDN:CAA35113.1; PID:g47793  
 C:Superfamily: nitroreductase  
 C:Keywords: oxidoreductase

Alignment Scores:  
 Pred. No.: 1.76e-10 Length: 217  
 Score: 196.50 Matches: 52  
 Percent Similarity: 48.35% Conservative: 36  
 Best Local Similarity: 28.57% Mismatches: 55  
 Query Match: 12.60% Indels: 39  
 DB: 1 Gaps: 5

US-09-647-661-1 (1-900) x S08397 (1-217)

QY 193 GAGCGCCATCTTCAGAGATCTTTGACAGCCATTATGAGTTTCTAGTGAAGATTAGAA 252  
 Db 9 GlnArgTyrSerThrLysAlaPheAspProSerLysLysLeuThrAlaGluAlaAsp 28  
 QY 253 GAATTCGCTGAATCGCCAGGCTATCGCAAGCTCTTACACAGCGCATCGCATTTT 312  
 Db 29 LysIleLysThrLeuLeuGlnTyrSerProSerSerThrAsnSerGlnProTrpHisPhe.48  
 QY 313 GTGATGGTTACTAATAGGATTTAAACAAACAAATTCGAGTGCACAGC----- 360  
 Db 49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnTyr 68  
 QY 361 TACTTTAATGAAGAAATGATTAAGCGCTTCACGCTTAATGGTGGTATCTCTTTAAGA 420  
 Db 69 ThrPheAsnGluArgLysMetLeuAspAlaSerHisValValPheCysAla----- 86  
 QY 421 CCTAGCGAGTTGTACCACAGCGCATTTACATGCAAAACCTTTTACC CGGAGTCTTATAAG 480  
 Db 87 -----LysThrAlaMetAspAlaTrpLeu 95  
 QY 481 GTTAGAGTGATC-----CCTTCTTTT 501  
 Db 96 GluArgValValAspGlnGluAspAlaAspGlyArgPheAlaThrProGluAlaLysAla 115  
 QY 502 GCTCAATGCTTGGCGTGGAGATTCACACAGCATGCAAGATTA----- 546  
 Db 116 AlaAsnAspLysGlyArgPhePheAlaAspMetHisArgValSerLeuLysAspAsp 135  
 QY 547 GAAAGCTATATTTAGACCAATGCTATATCGCTGTGGGGCAAAATTTGCATGGCGGTGAGC 606  
 Db 136 HisGlnTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeuGlyValAla 155  
 QY 607 TTAATGGATTGGATTCATTTATGGAGCCTTTCATCCCTTTAAAGTGGGTGAAGTT 666  
 Db 156 AlaMetGlyLeuAspAlaValProIleGluGlyPheAsp-----AlaGluVal 171

QY 667 TTAGAA 672  
 Db 172 LeuAsp 173  
 RESULT 15  
 AD0573  
 oxygen-insensitive NAD(P)H nitroreductase [imported] - Salmonella enterica subsp. enter  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AD0573  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher  
 th, T.; Conington, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
 , S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AD0573  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-217 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD05054.1; PID:g16501830; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY0620  
 C:Superfamily: nitroreductase

Alignment Scores:  
 Pred. No.: 2.72e-10 Length: 217  
 Score: 194.50 Matches: 51  
 Percent Similarity: 48.35% Conservative: 37  
 Best Local Similarity: 28.02% Mismatches: 55  
 Query Match: 12.48% Indels: 39  
 DB: 2 Gaps: 5

US-09-647-661-1 (1-900) x AD0573 (1-217)

QY 193 GAGCGCCATCTTCAGAGATCTTTGACAGCCATTATGAGTTTCTAGTGAAGATTAGAA 252  
 Db 9 LysArgTyrSerThrLysAlaPheAspProSerLysLysLeuThrAlaGluAlaAsp 28  
 QY 253 GAATTCGCTGAATCGCCAGGCTATCGCAAGCTCTTACACAGCGCATCGCATTTT 312  
 Db 29 LysValLysThrLeuLeuGlnTyrSerProSerSerThrAsnSerGlnProTrpHisPhe 48  
 QY 313 GTGATGGTTACTAATAGGATTTAAACAAACAAATTCGAGTGCACAGC----- 360  
 Db 49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnTyr 68  
 QY 361 TACTTTAATGAAGAAATGATTAAGCGCTTCACGCTTAATGGTGGTATCTCTTTAAGA 420  
 Db 69 ThrPheAsnGluArgLysMetLeuAspAlaSerHisValValPheCysAla----- 86  
 QY 421 CCTAGCGAGTTGTACCACAGCGCATTTACATGCAAAACCTTTTACC CGGAGTCTTATAAG 480  
 Db 87 -----LysThrAlaMetAspAlaTrpLeu 95  
 QY 481 GTTAGAGTGATC-----CCTTCTTTT 501  
 Db 96 GluArgValValAspGlnGluAspAlaAspGlyArgPheAlaThrProGluAlaLysAla 115  
 QY 502 GCTCAATGCTTGGCGTGGAGATTCACACAGCATGCAAGATTA----- 546  
 Db 116 AlaAsnAspLysGlyArgPhePheAlaAspMetHisArgValSerLeuLysAspAsp 135  
 QY 547 GAAAGCTATATTTAGACCAATGCTATATCGCTGTGGGGCAAAATTTGCATGGCGGTGAGC 606  
 Db 136 HisGlnTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeuGlyValAla 155  
 QY 607 TTAATGGATTGGATTCATTTATGGAGCCTTTCATCCCTTTAAAGTGGGTGAAGTT 666  
 Db 156 AlaMetGlyLeuAspAlaValProIleGluGlyPheAsp-----AlaGluVal 171

Qy 667 TTAGAA 672  
|||:::  
Db 172 LeuAsp 173

Search completed: April 20, 2004, 19:29:01  
Job time : 28.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 20, 2004, 19:12:57 ; Search time 15 Seconds

(without alignments)  
6248.419 Million cell updates/sec

Title: US-09-647-661-1

Perfect score: 1559

Sequence: 1 tgcagattttacagagc.....gggggtttttcaagggtttc 900

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=cgn2\_1/USPTO\_spool\_p/US09647661/runat\_20042004\_135347\_2651/app\_query.fasta\_1.1095  
-DB=SwissProt\_42 -QFMT=fastan -SUPFIX=rs -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=Blosum62 -TRANS=numat40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09647661@cgn 1 16 @runat\_20042004\_135347\_2651 -NCPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246	15.8	284	1 LGT_HELPY	O25609 helicobacte
2	229	14.7	283	1 LGT_HELPY	O82k06 helicobacte
3	223.5	14.3	221	1 YFPO_BACSU	O34475 bacillus su
4	218.5	14.0	220	1 YC78_HAEIN	Q57431 haemophilus
5	198.5	12.7	217	1 NFNB_ENTCL	Q01234 enterobacte
6	196.5	12.6	217	1 NFNB_SALTY	P15888 salmonella
7	188.5	12.1	217	1 NFNB_ECOLI	P38489 escherichia
8	185.5	11.9	201	1 V12C_BACSU	P81102 bacillus su
9	172.5	11.1	209	1 YGCI_BACSU	P95707 bacillus su
10	169.5	10.9	210	1 DRGA_SYNT3	Q55233 synchocyst
11	154	9.9	217	1 FRAI_VIBET	P46072 vibrio fisc
12	153	9.8	174	1 YM67_ARCFU	O28017 archaeoglob
13	137.5	8.8	206	1 YDFN_BACSU	P96692 bacillus su
14	136.5	8.8	205	1 NOX_TRETH	Q60049 thermus the
15	126	8.1	191	1 Y120_METTH	O26223 methanobact
16	124	8.0	172	1 Y226_ARCFU	O30013 archaeoglob
17	103.5	6.6	359	1 YK45_CABEL	P34273 caenorhabdi
18	97.5	6.3	274	1 DAPD_BUCAL	P57323 buchnera ap

19	91	5.8	1125	1 MFD_BORBU	O51568 borrelia bu
20	90.5	5.8	281	1 LGT_BUCAL	P57514 buchnera ap
21	90.5	5.8	353	1 Y772_METIA	Q58182 methanococ
22	89	5.7	1001	1 MFD_HELPY	Q2zj57 helicobacte
23	88	5.6	184	1 YDJA_HAEIN	P45244 haemophilus
24	87.5	5.6	202	1 YZEL_STRCO	Q95bvo streptomyce
25	87.5	5.6	273	1 DAPD_BUCAP	O85290 buchnera ap
26	87.5	5.6	291	1 LGT_ECOLI	P37149 escherichia
27	87.5	5.6	901	1 MALT_ECOLI	P06993 escherichia
28	86.5	5.5	196	1 Y096_AGRTS	P58792 agrobacteri
29	86.5	5.5	999	1 MFD_HELPY	O26066 helicobacte
30	86	5.7	315	1 OSAY_HUMAN	Q8g22 homo sapien
31	85.5	5.5	459	1 MURD_LACPL	Q88v80 lactobacill
32	85	5.5	484	1 SYE_MYCGE	P47700 mycoplasma
33	85	5.5	901	1 R8L3_ARATH	Q9fjb5 arabidopsis
34	84.5	5.4	245	1 KDSB_FUSNN	O8rfa8 fusobacteri
35	84	5.4	452	1 DHAP_RAT	P11883 rattus norv
36	84	5.4	1102	1 MISC_CHICK	P39616 gallus gall
37	83.5	5.4	207	1 RPOA_EUGAN	Q8s194 euglena ana
38	83.5	5.4	1581	1 ACC8_CRICR	Q09427 cricetus cr
39	83	5.3	249	1 NFPA_BACSV	P39605 bacillus su
40	83	5.3	472	1 PAP1_FOWPV	Q9j5b8 fowlpox vir
41	83	5.3	535	1 Y295_STR3A	O86762 streptococc
42	83	5.3	888	1 LONI_ARATH	O64948 arabidopsis
43	83	5.3	1276	1 BXA_CLOBO	P19321 clostridium
44	82.5	5.3	751	1 TALA_POVHA	P03075 hamster pol
45	82.5	5.3	1161	1 BM2K_HUMAN	Q9msy1 homo sapien

#### ALIGNMENTS

RESULT 1  
LGT\_HELPY  
ID LGT\_HELPY STANDARD; PRT; 284 AA.  
AC O25609; O30584; O30586;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Prolipoprotein diacylglycerol transferase (EC 2.4.99.-).  
GN LGT OR HP0955.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=26695 / ATCC 700392;  
RX MEDLINE=97394467; PubMed=9252185;  
RA Tomb J.-F., White O., Karlavage A.R., Clayton R.A., Sutton G.G.,  
RA Fleisemann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,  
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
RA Venter J.C.;  
RA "The complete genome sequence of the gastric pathogen Helicobacter  
pylori.";  
RL Nature 388:539-547(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HP439, and HP500;  
RX Goodwin A., Berg D.E., Hoffman P.S.;  
RA Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RL Submitter Transfers the N-acyl diglyceride group on what will  
CC become the N-terminal cysteine of membrane lipoproteins (By  
CC similarity).  
CC -1- PATHWAY: Lipoproteins biosynthesis; first step.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (Potential).  
CC -1- SIMILARITY: Belongs to the lgt family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB000604; AAC07998.1; -  
 CC EMBL; AF012552; AAC46348.1; -  
 CC EMBL; AF012553; AAC46350.1; -  
 CC PIR; C64639; C64639.  
 CC TIGR; HP0955; -  
 CC InterPro: IPR001640; LGT.  
 CC Pfam; PF01790; LGT; 1.  
 CC ProDom; PD005412; LGT; 1.  
 CC TIGRFAMs; TIGR00544; LGT; 1.  
 CC PROSITE; PS01311; LGT; 1.  
 CC TRANSFERASE; Transmembrane; Inner membrane; Complete proteome.  
 CC FT TRANSMEM 14 34  
 CC FT TRANSMEM 62 82  
 CC FT TRANSMEM 106 126  
 CC FT TRANSMEM 136 156  
 CC FT TRANSMEM 190 210  
 CC FT TRANSMEM 218 238  
 CC FT TRANSMEM 252 272  
 CC FT TRANSMEM 284 326  
 CC FT TRANSMEM 326 364  
 CC FT TRANSMEM 364 402  
 CC FT TRANSMEM 402 440  
 CC FT TRANSMEM 440 478  
 CC FT TRANSMEM 478 516  
 CC FT TRANSMEM 516 554  
 CC FT TRANSMEM 554 592  
 CC FT TRANSMEM 592 630  
 CC FT TRANSMEM 630 668  
 CC FT TRANSMEM 668 706  
 CC FT TRANSMEM 706 744  
 CC FT TRANSMEM 744 782  
 CC FT TRANSMEM 782 820  
 CC FT TRANSMEM 820 858  
 CC FT TRANSMEM 858 896  
 CC FT TRANSMEM 896 934  
 CC FT TRANSMEM 934 972  
 CC FT TRANSMEM 972 1010  
 CC FT TRANSMEM 1010 1048  
 CC FT TRANSMEM 1048 1086  
 CC FT TRANSMEM 1086 1124  
 CC FT TRANSMEM 1124 1162  
 CC FT TRANSMEM 1162 1200  
 CC FT TRANSMEM 1200 1238  
 CC FT TRANSMEM 1238 1276  
 CC FT TRANSMEM 1276 1314  
 CC FT TRANSMEM 1314 1352  
 CC FT TRANSMEM 1352 1390  
 CC FT TRANSMEM 1390 1428  
 CC FT TRANSMEM 1428 1466  
 CC FT TRANSMEM 1466 1504  
 CC FT TRANSMEM 1504 1542  
 CC FT TRANSMEM 1542 1580  
 CC FT TRANSMEM 1580 1618  
 CC FT TRANSMEM 1618 1656  
 CC FT TRANSMEM 1656 1694  
 CC FT TRANSMEM 1694 1732  
 CC FT TRANSMEM 1732 1770  
 CC FT TRANSMEM 1770 1808  
 CC FT TRANSMEM 1808 1846  
 CC FT TRANSMEM 1846 1884  
 CC FT TRANSMEM 1884 1922  
 CC FT TRANSMEM 1922 1960  
 CC FT TRANSMEM 1960 1998  
 CC FT TRANSMEM 1998 2036  
 CC FT TRANSMEM 2036 2074  
 CC FT TRANSMEM 2074 2112  
 CC FT TRANSMEM 2112 2150  
 CC FT TRANSMEM 2150 2188  
 CC FT TRANSMEM 2188 2226  
 CC FT TRANSMEM 2226 2264  
 CC FT TRANSMEM 2264 2302  
 CC FT TRANSMEM 2302 2340  
 CC FT TRANSMEM 2340 2378  
 CC FT TRANSMEM 2378 2416  
 CC FT TRANSMEM 2416 2454  
 CC FT TRANSMEM 2454 2492  
 CC FT TRANSMEM 2492 2530  
 CC FT TRANSMEM 2530 2568  
 CC FT TRANSMEM 2568 2606  
 CC FT TRANSMEM 2606 2644  
 CC FT TRANSMEM 2644 2682  
 CC FT TRANSMEM 2682 2720  
 CC FT TRANSMEM 2720 2758  
 CC FT TRANSMEM 2758 2796  
 CC FT TRANSMEM 2796 2834  
 CC FT TRANSMEM 2834 2872  
 CC FT TRANSMEM 2872 2910  
 CC FT TRANSMEM 2910 2948  
 CC FT TRANSMEM 2948 2986  
 CC FT TRANSMEM 2986 3024  
 CC FT TRANSMEM 3024 3062  
 CC FT TRANSMEM 3062 3100  
 CC FT TRANSMEM 3100 3138  
 CC FT TRANSMEM 3138 3176  
 CC FT TRANSMEM 3176 3214  
 CC FT TRANSMEM 3214 3252  
 CC FT TRANSMEM 3252 3290  
 CC FT TRANSMEM 3290 3328  
 CC FT TRANSMEM 3328 3366  
 CC FT TRANSMEM 3366 3404  
 CC FT TRANSMEM 3404 3442  
 CC FT TRANSMEM 3442 3480  
 CC FT TRANSMEM 3480 3518  
 CC FT TRANSMEM 3518 3556  
 CC FT TRANSMEM 3556 3594  
 CC FT TRANSMEM 3594 3632  
 CC FT TRANSMEM 3632 3670  
 CC FT TRANSMEM 3670 3708  
 CC FT TRANSMEM 3708 3746  
 CC FT TRANSMEM 3746 3784  
 CC FT TRANSMEM 3784 3822  
 CC FT TRANSMEM 3822 3860  
 CC FT TRANSMEM 3860 3898  
 CC FT TRANSMEM 3898 3936  
 CC FT TRANSMEM 3936 3974  
 CC FT TRANSMEM 3974 4012  
 CC FT TRANSMEM 4012 4050  
 CC FT TRANSMEM 4050 4088  
 CC FT TRANSMEM 4088 4126  
 CC FT TRANSMEM 4126 4164  
 CC FT TRANSMEM 4164 4202  
 CC FT TRANSMEM 4202 4240  
 CC FT TRANSMEM 4240 4278  
 CC FT TRANSMEM 4278 4316  
 CC FT TRANSMEM 4316 4354  
 CC FT TRANSMEM 4354 4392  
 CC FT TRANSMEM 4392 4430  
 CC FT TRANSMEM 4430 4468  
 CC FT TRANSMEM 4468 4506  
 CC FT TRANSMEM 4506 4544  
 CC FT TRANSMEM 4544 4582  
 CC FT TRANSMEM 4582 4620  
 CC FT TRANSMEM 4620 4658  
 CC FT TRANSMEM 4658 4696  
 CC FT TRANSMEM 4696 4734  
 CC FT TRANSMEM 4734 4772  
 CC FT TRANSMEM 4772 4810  
 CC FT TRANSMEM 4810 4848  
 CC FT TRANSMEM 4848 4886  
 CC FT TRANSMEM 4886 4924  
 CC FT TRANSMEM 4924 4962  
 CC FT TRANSMEM 4962 5000  
 CC FT TRANSMEM 5000 5038  
 CC FT TRANSMEM 5038 5076  
 CC FT TRANSMEM 5076 5114  
 CC FT TRANSMEM 5114 5152  
 CC FT TRANSMEM 5152 5190  
 CC FT TRANSMEM 5190 5228  
 CC FT TRANSMEM 5228 5266  
 CC FT TRANSMEM 5266 5304  
 CC FT TRANSMEM 5304 5342  
 CC FT TRANSMEM 5342 5380  
 CC FT TRANSMEM 5380 5418  
 CC FT TRANSMEM 5418 5456  
 CC FT TRANSMEM 5456 5494  
 CC FT TRANSMEM 5494 5532  
 CC FT TRANSMEM 5532 5570  
 CC FT TRANSMEM 5570 5608  
 CC FT TRANSMEM 5608 5646  
 CC FT TRANSMEM 5646 5684  
 CC FT TRANSMEM 5684 5722  
 CC FT TRANSMEM 5722 5760  
 CC FT TRANSMEM 5760 5798  
 CC FT TRANSMEM 5798 5836  
 CC FT TRANSMEM 5836 5874  
 CC FT TRANSMEM 5874 5912  
 CC FT TRANSMEM 5912 5950  
 CC FT TRANSMEM 5950 5988  
 CC FT TRANSMEM 5988 6026  
 CC FT TRANSMEM 6026 6064  
 CC FT TRANSMEM 6064 6102  
 CC FT TRANSMEM 6102 6140  
 CC FT TRANSMEM 6140 6178  
 CC FT TRANSMEM 6178 6216  
 CC FT TRANSMEM 6216 6254  
 CC FT TRANSMEM 6254 6292  
 CC FT TRANSMEM 6292 6330  
 CC FT TRANSMEM 6330 6368  
 CC FT TRANSMEM 6368 6406  
 CC FT TRANSMEM 6406 6444  
 CC FT TRANSMEM 6444 6482  
 CC FT TRANSMEM 6482 6520  
 CC FT TRANSMEM 6520 6558  
 CC FT TRANSMEM 6558 6596  
 CC FT TRANSMEM 6596 6634  
 CC FT TRANSMEM 6634 6672  
 CC FT TRANSMEM 6672 6710  
 CC FT TRANSMEM 6710 6748  
 CC FT TRANSMEM 6748 6786  
 CC FT TRANSMEM 6786 6824  
 CC FT TRANSMEM 6824 6862  
 CC FT TRANSMEM 6862 6900  
 CC FT TRANSMEM 6900 6938  
 CC FT TRANSMEM 6938 6976  
 CC FT TRANSMEM 6976 7014  
 CC FT TRANSMEM 7014 7052  
 CC FT TRANSMEM 7052 7090  
 CC FT TRANSMEM 7090 7128  
 CC FT TRANSMEM 7128 7166  
 CC FT TRANSMEM 7166 7204  
 CC FT TRANSMEM 7204 7242  
 CC FT TRANSMEM 7242 7280  
 CC FT TRANSMEM 7280 7318  
 CC FT TRANSMEM 7318 7356  
 CC FT TRANSMEM 7356 7394  
 CC FT TRANSMEM 7394 7432  
 CC FT TRANSMEM 7432 7470  
 CC FT TRANSMEM 7470 7508  
 CC FT TRANSMEM 7508 7546  
 CC FT TRANSMEM 7546 7584  
 CC FT TRANSMEM 7584 7622  
 CC FT TRANSMEM 7622 7660  
 CC FT TRANSMEM 7660 7698  
 CC FT TRANSMEM 7698 7736  
 CC FT TRANSMEM 7736 7774  
 CC FT TRANSMEM 7774 7812  
 CC FT TRANSMEM 7812 7850  
 CC FT TRANSMEM 7850 7888  
 CC FT TRANSMEM 7888 7926  
 CC FT TRANSMEM 7926 7964  
 CC FT TRANSMEM 7964 8002  
 CC FT TRANSMEM 8002 8040  
 CC FT TRANSMEM 8040 8078  
 CC FT TRANSMEM 8078 8116  
 CC FT TRANSMEM 8116 8154  
 CC FT TRANSMEM 8154 8192  
 CC FT TRANSMEM 8192 8230  
 CC FT TRANSMEM 8230 8268  
 CC FT TRANSMEM 8268 8306  
 CC FT TRANSMEM 8306 8344  
 CC FT TRANSMEM 8344 8382  
 CC FT TRANSMEM 8382 8420  
 CC FT TRANSMEM 8420 8458  
 CC FT TRANSMEM 8458 8496  
 CC FT TRANSMEM 8496 8534  
 CC FT TRANSMEM 8534 8572  
 CC FT TRANSMEM 8572 8610  
 CC FT TRANSMEM 8610 8648  
 CC FT TRANSMEM 8648 8686  
 CC FT TRANSMEM 8686 8724  
 CC FT TRANSMEM 8724 8762  
 CC FT TRANSMEM 8762 8800  
 CC FT TRANSMEM 8800 8838  
 CC FT TRANSMEM 8838 8876  
 CC FT TRANSMEM 8876 8914  
 CC FT TRANSMEM 8914 8952  
 CC FT TRANSMEM 8952 8990  
 CC FT TRANSMEM 8990 9028  
 CC FT TRANSMEM 9028 9066  
 CC FT TRANSMEM 9066 9104  
 CC FT TRANSMEM 9104 9142  
 CC FT TRANSMEM 9142 9180  
 CC FT TRANSMEM 9180 9218  
 CC FT TRANSMEM 9218 9256  
 CC FT TRANSMEM 9256 9294  
 CC FT TRANSMEM 9294 9332  
 CC FT TRANSMEM 9332 9370  
 CC FT TRANSMEM 9370 9408  
 CC FT TRANSMEM 9408 9446  
 CC FT TRANSMEM 9446 9484  
 CC FT TRANSMEM 9484 9522  
 CC FT TRANSMEM 9522 9560  
 CC FT TRANSMEM 9560 9598  
 CC FT TRANSMEM 9598 9636  
 CC FT TRANSMEM 9636 9674  
 CC FT TRANSMEM 9674 9712  
 CC FT TRANSMEM 9712 9750  
 CC FT TRANSMEM 9750 9788  
 CC FT TRANSMEM 9788 9826  
 CC FT TRANSMEM 9826 9864  
 CC FT TRANSMEM 9864 9902  
 CC FT TRANSMEM 9902 9940  
 CC FT TRANSMEM 9940 9978  
 CC FT TRANSMEM 9978 10016  
 CC FT TRANSMEM 10016 10054  
 CC FT TRANSMEM 10054 10092  
 CC FT TRANSMEM 10092 10130  
 CC FT TRANSMEM 10130 10168  
 CC FT TRANSMEM 10168 10206  
 CC FT TRANSMEM 10206 10244  
 CC FT TRANSMEM 10244 10282  
 CC FT TRANSMEM 10282 10320  
 CC FT TRANSMEM 10320 10358  
 CC FT TRANSMEM 10358 10396  
 CC FT TRANSMEM 10396 10434  
 CC FT TRANSMEM 10434 10472  
 CC FT TRANSMEM 10472 10510  
 CC FT TRANSMEM 10510 10548  
 CC FT TRANSMEM 10548 10586  
 CC FT TRANSMEM 10586 10624  
 CC FT TRANSMEM 10624 10662  
 CC FT TRANSMEM 10662 10700  
 CC FT TRANSMEM 10700 10738  
 CC FT TRANSMEM 10738 10776  
 CC FT TRANSMEM 10776 10814  
 CC FT TRANSMEM 10814 10852  
 CC FT TRANSMEM 10852 10890  
 CC FT TRANSMEM 10890 10928  
 CC FT TRANSMEM 10928 10966  
 CC FT TRANSMEM 10966 11004  
 CC FT TRANSMEM 11004 11042  
 CC FT TRANSMEM 11042 11080  
 CC FT TRANSMEM 11080 11118  
 CC FT TRANSMEM 11118 11156  
 CC FT TRANSMEM 11156 11194  
 CC FT TRANSMEM 11194 11232  
 CC FT TRANSMEM 11232 11270  
 CC FT TRANSMEM 11270 11308  
 CC FT TRANSMEM 11308 11346  
 CC FT TRANSMEM 11346 11384  
 CC FT TRANSMEM 11384 11422  
 CC FT TRANSMEM 11422 11460  
 CC FT TRANSMEM 11460 11498  
 CC FT TRANSMEM 11498 11536  
 CC FT TRANSMEM 11536 11574  
 CC FT TRANSMEM 11574 11612  
 CC FT TRANSMEM 11612 11650  
 CC FT TRANSMEM 11650 11688  
 CC FT TRANSMEM 11688 11726  
 CC FT TRANSMEM 11726 11764  
 CC FT TRANSMEM 11764 11802  
 CC FT TRANSMEM 11802 11840  
 CC FT TRANSMEM 11840 11878  
 CC FT TRANSMEM 11878 11916  
 CC FT TRANSMEM 11916 11954  
 CC FT TRANSMEM 11954 11992  
 CC FT TRANSMEM 11992 12030  
 CC FT TRANSMEM 12030 12068  
 CC FT TRANSMEM 12068 12106  
 CC FT TRANSMEM 12106 12144  
 CC FT TRANSMEM 12144 12182  
 CC FT TRANSMEM 12182 12220  
 CC FT TRANSMEM 12220 12258  
 CC FT TRANSMEM 12258 12296  
 CC FT TRANSMEM 12296 12334  
 CC FT TRANSMEM 12334 12372  
 CC FT TRANSMEM 12372 12410  
 CC FT TRANSMEM 12410 12448  
 CC FT TRANSMEM 12448 12486  
 CC FT TRANSMEM 12486 12524  
 CC FT TRANSMEM 12524 12562  
 CC FT TRANSMEM 12562 12600  
 CC FT TRANSMEM 12600 12638  
 CC FT TRANSMEM 12638 12676  
 CC FT TRANSMEM 12676 12714  
 CC FT TRANSMEM 12714 12752  
 CC FT TRANSMEM 12752 12790  
 CC FT TRANSMEM 12790 12828  
 CC FT TRANSMEM 12828 12866  
 CC FT TRANSMEM 12866 12904  
 CC FT TRANSMEM 12904 12942  
 CC FT TRANSMEM 12942 12980  
 CC FT TRANSMEM 12980 13018  
 CC FT TRANSMEM 13018 13056  
 CC FT TRANSMEM 13056 13094  
 CC FT TRANSMEM 13094 13132  
 CC FT TRANSMEM 13132 13170  
 CC FT TRANSMEM 13170 13208  
 CC FT TRANSMEM 13208 13246  
 CC FT TRANSMEM 13246 13284  
 CC FT TRANSMEM 13284 13322  
 CC FT TRANSMEM 13322 13360  
 CC FT TRANSMEM 13360 13398  
 CC FT TRANSMEM 13398 13436  
 CC FT TRANSMEM 13436 13474  
 CC FT TRANSMEM 13474 13512  
 CC FT TRANSMEM 13512 13550  
 CC FT TRANSMEM 13550 13588  
 CC FT TRANSMEM 13588 13626  
 CC FT TRANSMEM 13626 13664  
 CC FT TRANSMEM 13664 13702  
 CC FT TRANSMEM 13702 13740  
 CC FT TRANSMEM 13740 13778  
 CC FT TRANSMEM 13778 13816  
 CC FT TRANSMEM 13816 13854  
 CC FT TRANSMEM 13854 13892  
 CC FT TRANSMEM 13892 13930  
 CC FT TRANSMEM 13930 13968  
 CC FT TRANSMEM 13968 14006  
 CC FT TRANSMEM 14006 14044  
 CC FT TRANSMEM 14044 14082  
 CC FT TRANSMEM 14082 14120  
 CC FT TRANSMEM 14120 14158  
 CC FT TRANSMEM 14158 14196  
 CC FT TRANSMEM 14196 14234  
 CC FT TRANSMEM 14234 14272  
 CC FT TRANSMEM 14272 14310  
 CC FT TRANSMEM 14310 14348  
 CC FT TRANSMEM 14348 14386  
 CC FT TRANSMEM 14386 14424  
 CC FT TRANSMEM 14424 14462  
 CC FT TRANSMEM 14462 14500  
 CC FT TRANSMEM 14500 14538  
 CC FT TRANSMEM 14538 14576  
 CC FT TRANSMEM 14576 14614  
 CC FT TRANSMEM 14614 14652  
 CC FT TRANSMEM 14652 14690  
 CC FT TRANSMEM 14690 14728  
 CC FT TRANSMEM 14728 14766  
 CC FT TRANSMEM 14766 14804  
 CC FT TRANSMEM 14804 14842  
 CC FT TRANSMEM 14842 14880  
 CC FT TRANSMEM 14880 14918  
 CC FT TRANSMEM 14918 14956  
 CC FT TRANSMEM 14956 14994  
 CC FT TRANSMEM 14994 15032  
 CC FT TRANSMEM 15032 15070  
 CC FT TRANSMEM 15070 15108  
 CC FT TRANSMEM 15108 15146  
 CC FT TRANSMEM 15146 15184  
 CC FT TRANSMEM 15184 15222  
 CC FT TRANSMEM 15222 15260  
 CC FT TRANSMEM 15260 15298  
 CC FT TRANSMEM 15298 15336  
 CC FT TRANSMEM 15336 15374  
 CC FT TRANSMEM 15374 15412  
 CC FT TRANSMEM 15412 15450  
 CC FT TRANSMEM 15450 15488  
 CC FT TRANSMEM 15488 15526  
 CC FT TRANSMEM 15526 15564  
 CC FT TRANSMEM 15564 15602  
 CC FT TRANSMEM 15602 15640  
 CC FT TRANSMEM 15640 15678  
 CC FT TRANSMEM 15678 15716  
 CC FT TRANSMEM 15716 15754  
 CC FT TRANSMEM 15754 15792  
 CC FT TRANSMEM 15792 15830  
 CC FT TRANSMEM 15830 15868  
 CC FT TRANSMEM 15868 15906  
 CC FT TRANSMEM 15906 15944  
 CC FT TRANSMEM 15944 15982  
 CC FT TRANSMEM 15982 16020  
 CC FT TRANSMEM 16020 16058  
 CC FT TRANSMEM 16058 16096  
 CC FT TRANSMEM 16096 16134  
 CC FT TRANSMEM 16134 16172  
 CC FT TRANSMEM 16172 16210  
 CC FT TRANSMEM 16210 16248  
 CC FT TRANSMEM 16248 16286  
 CC FT TRANSMEM 16286 16324  
 CC FT TRANSMEM 16324 16362  
 CC FT TRANSMEM 16362 16400  
 CC FT TRANSMEM 16400 16438  
 CC FT TRANSMEM 16438 16476  
 CC FT TRANSMEM 16476 16514  
 CC FT TRANSMEM 16514 16552  
 CC FT TRANSMEM 16552 16590  
 CC FT TRANSMEM 16590 16628  
 CC FT TRANSMEM 16628 16666  
 CC FT TRANSMEM 16666 16704  
 CC FT TRANSMEM 16704 16742  
 CC FT TRANSMEM 16742 16780  
 CC FT TRANSMEM 16780 16818  
 CC FT TRANSMEM 16818 16856  
 CC FT TRANSMEM 16856 16894  
 CC FT TRANSMEM 16894 16932  
 CC FT TRANSMEM 16932 16970  
 CC FT TRANSMEM 16970 17008  
 CC FT TRANSMEM 17008 17046  
 CC FT TRANSMEM 17046 17084  
 CC FT TRANSMEM 17084 17122  
 CC FT TRANSMEM 17122 17160  
 CC FT TRANSMEM 17160 17198  
 CC FT TRANSMEM 17198 17236  
 CC FT TRANSMEM 17236 17274  
 CC FT TRANSMEM 17274 17312  
 CC FT TRANSMEM 17312 17350  
 CC FT TRANSMEM 17350 17388  
 CC FT TRANSMEM 17388 17426  
 CC FT TRANSMEM 17426 17464  
 CC FT TRANSMEM 17464 17502  
 CC FT TRANSMEM 17502 17540  
 CC FT TRANSMEM 17540 17578  
 CC FT TRANSMEM 17578 17616  
 CC FT TRANSMEM 17616 17654  
 CC FT TRANSMEM 17654 17692  
 CC FT TRANSMEM 17692 17730  
 CC FT TRANSMEM 17730 17768  
 CC FT TRANSMEM 17768 17806  
 CC FT TRANSMEM 17806 17844  
 CC FT TRANSMEM 17844 17882  
 CC FT TRANSMEM 17882 17920  
 CC FT TRANSMEM 17920 17958  
 CC FT TRANSMEM 17958 17996  
 CC FT TRANSMEM 17996 18034  
 CC FT TRANSMEM 18034 18072  
 CC FT TRANSMEM 18072 18110  
 CC FT TRANSMEM 18110 18148  
 CC FT TRANSMEM 18148 18186  
 CC FT TRANSMEM 18186 18224  
 CC FT TRANSMEM 18224 18262  
 CC FT TRANSMEM 18262 18300  
 CC FT TRANSMEM 18300 18338  
 CC FT TRANSMEM 18338 18376  
 CC FT TRANSMEM 18376 18414  
 CC FT TRANSMEM 18414 18452  
 CC FT TRANSMEM 18452 18490  
 CC FT TRANSMEM 18490 18528  
 CC FT TRANSMEM 18528 18566  
 CC FT TRANSMEM 18566 18604  
 CC FT TRANSMEM 18604 18642  
 CC FT TRANSMEM 18642 18680  
 CC FT TRANSMEM 18680 18718  
 CC FT TRANSMEM 18718 18756  
 CC FT TRANSMEM 18756 18794  
 CC FT TRANSMEM 18794 18832  
 CC FT TRANSMEM 18832 18870  
 CC FT TRANSMEM 18870 18908  
 CC FT TRANSMEM 18908 18946  
 CC FT TRANSMEM 18946 18984  
 CC FT TRANSMEM 18984 19022  
 CC FT TRANSMEM 19022 19060  
 CC FT TRANSMEM 19060 19098  
 CC FT TRANSMEM 19098 19136  
 CC FT TRANSMEM 19136 19174  
 CC FT TRANSMEM 19174 19212  
 CC FT TRANSMEM 19212 19250  
 CC FT TRANSMEM 19250 19288  
 CC FT TRANSMEM 19288 19326  
 CC FT TRANSMEM 19326 19364  
 CC FT TRANSMEM 19364 19402  
 CC FT TRANSMEM 19402 19440  
 CC FT TRANSMEM 19440 19478  
 CC FT TRANSMEM 19478 19516  
 CC FT TRANSMEM 19516 19554  
 CC FT TRANSMEM 19554 19592  
 CC FT TRANSMEM 19592 19630  
 CC FT TRANSMEM 19630 19668  
 CC FT TRANSMEM 19668 19706  
 CC FT TRANSMEM 19706 19744  
 CC FT TRANSMEM 19744 19782  
 CC FT TRANSMEM 19782 19820  
 CC FT TRANSMEM 19820 19858  
 CC FT TRANSMEM 19858 19896  
 CC FT TRANSMEM 19896 19934  
 CC FT TRANSMEM 19934 19972  
 CC FT TRANSMEM 19972 20010  
 CC FT TRANSMEM 20010 20048  
 CC FT TRANSMEM 20048 20086  
 CC FT TRANSMEM 20086 20124  
 CC FT TRANSMEM 20124 20162  
 CC FT TRANSMEM 20162 20200  
 CC FT TRANSMEM 20200 20238  
 CC FT TRANSMEM 20238 20276  
 CC FT TRANSMEM 20276 20314  
 CC FT TRANSMEM 20314 20352  
 CC FT TRANSMEM 20352 20390  
 CC FT TRANSMEM 20390 20428  
 CC FT TRANSMEM 20428 20466  
 CC FT TRANSMEM 20466 20504  
 CC FT TRANSMEM 20504 20542  
 CC FT TRANSMEM 20542 20580  
 CC FT TRANSMEM 20580 20618  
 CC FT TRANSMEM 20618 20656  
 CC FT TRANSMEM 20656 20694  
 CC FT TRANSMEM 20694 20732  
 CC FT TRANSMEM 20732 20770  
 CC FT TRANSMEM 20770 20808  
 CC FT TRANSMEM 20808 20846  
 CC FT TRANSMEM 20846 20884  
 CC FT TRANSMEM 20884 20922  
 CC FT TRANSMEM 20922 20960  
 CC FT TRANSMEM 20960 21000  
 CC FT TRANSMEM 21000 21040  
 CC FT TRANSMEM 21040 21080  
 CC FT TRANSMEM 21080 21120  
 CC FT TRANSMEM 21120 21160  
 CC FT TRANSMEM 21160 21200  
 CC FT TRANSMEM 21200 21240  
 CC FT TRANSMEM 21240 21280  
 CC FT TRANSMEM 21280 21320  
 CC FT TRANSMEM 21320 21360  
 CC FT TRANSMEM 21360 21400  
 CC FT TRANSMEM 21400 21440  
 CC FT TRANSMEM 21440 21480  
 CC FT TRANSMEM 21480 21520  
 CC FT TRANSMEM 21520 21560  
 CC FT TRANSMEM 21560 21600  
 CC FT TRANSMEM 21600 21640  
 CC FT TRANSMEM 21640 21680  
 CC FT TRANSMEM 21680 21720  
 CC FT TRANSMEM 21720 21760  
 CC FT TRANSMEM 21760 21800  
 CC FT TRANSMEM 21800 21840  
 CC FT TRANSMEM 21840 21880  
 CC FT TRANSMEM 21880 21920  
 CC FT TRANSMEM 21920 21960  
 CC FT TRANSMEM 21960 22000  
 CC FT TRANSMEM 22000 22040  
 CC FT TRANSMEM 22040 22080  
 CC FT TRANSMEM 22080 22120  
 CC FT TRANSMEM 22120 22160  
 CC FT TRANSMEM 22160 22200  
 CC FT TRANSMEM 22200 22240  
 CC FT TRANSMEM 22240 22280  
 CC FT TRANSMEM 22280 22320  
 CC FT TRANSMEM 22320 22360  
 CC FT TRANSMEM 22360 22400  
 CC FT TRANSMEM 22400 22440  
 CC FT TRANSMEM 22440 22480  
 CC FT TRANSMEM 22



Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=AC327;  
 RC Sekiguchi J., Yamamoto H., Uchiyama S., Fajar A.;  
 RC "Nucleotide sequence analysis of B. subtilis chromosome in 74 degree  
 RC region."; submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RC [2]  
 RN SEQUENCE FROM N.A.  
 RN STRAIN=168;  
 RN MEDLINE=98044033; PubMed=9384377;  
 RN Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RN Azevedo V., Bortner M.G., Bessieres P., Bolotin A., Borchert S.,  
 RN Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RN Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RN Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RN Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RN Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RN Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RN Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RN Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RN Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RN Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RN Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RN Kurita K., Lapidus A., Lardinouis S., Lauber J., Lazarevic V.,  
 RN Lee S.M., Levine A., Liu H., Masuda S., Maul C., Medigue C.,  
 RN Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RN Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RN Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RN Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RN Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RN Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RN Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RN Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RN Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RN Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RN Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,  
 RN Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RN Yoshida K., Yoshikawa H.P., Zumstein E., Yoshikawa H., Danchin A.,  
 RN "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RN subtilis";  
 RN Nature 390:249-256(1997).  
 RL - COFACTOR: FMN (Potential).  
 CC - SIMILARITY: Belongs to the nitroreductase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions in no way  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D83967; BAA23405.1; -  
 CC EMBL; Z99108; CAB2612.1; -  
 CC PIR; B69809; B69809.  
 CC Subtilist; BG12931; yfko.  
 CC InterPro; IPR000415; Nitroreductase.  
 CC Pfam; PF00881; Nitroreductase; 1.  
 CC Hypothetical protein; Oxidoreductase; NAD; NADP; Flavoprotein; FMN;  
 CC Complete proteome.  
 FT NP BIND 157 162 NAD OR NADP (BY SIMILARITY).  
 SQ SEQUENCE 221 AA; 25628 MW; 3D911EAC3402F599 CRC64;  
 Alignment Scores:  
 Pred. No.: 2,01e-13 Length: 221  
 Score: 223.50 Matches: 57  
 Percent Similarity: 49.09% Conservative: 51  
 Best Local Similarity: 25.91% Mismatches: 21  
 Query Match: 14,34% Indels: 21

Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.",  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=20137488; PubMed=10675023;  
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
 Gray C., Fountoulakis M.,  
 RA "Two-dimensional map of the proteome of Haemophilus influenzae";  
 RL Electrophoresis 21:411-429(2000).  
 CC -1- COPACTOR: FMN (Potential).  
 CC -1- SIMILARITY: Belongs to the nitroreductase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U32807; AAC22926.1; -  
 DR FIR; B64114; B64114.  
 DR HSP; P46072; 1VFR.  
 DR TIGR; H11278; -  
 DR InterPro: IPR000415; Nitroreductase.  
 DR Pfam; PF00881; Nitroreductase, 1.  
 DR Oxidoreductase; NAD; NADP; Flavoprotein; FMN; Complete proteome.  
 FT NP\_BIND 155 160 NAD OR NADP (BY SIMILARITY).  
 SQ SEQUENCE 220 AA; 25189 MW; AB0D804601BC975E CRC64;  
 Alignment Scores:  
 Pred. No.: 6.03e-13 Length: 220  
 Score: 218.50 Matches: 59  
 Percent Similarity: 48.89% Conservative: 51  
 Best Local Similarity: 26.22% Mismatches: 94  
 Query Match: 14.02% Indels: 21  
 DB: 1 Gaps: 4  
 US-09-647-661-1 (1-900) x YC78\_HABIN (1-220)  
 QY 151 ATGAATTTTGGATCAGGAAAAGAACACTATTAAACGAGCGCCATTCTTGCAG 210  
 DB 1 MetThrGlnLeuThrArgGluGlnValLeuGluLeuPheHisGlnArgSerThrArg 20  
 QY 211 ATGTTTACACGCCATTATGATTTCTAGTGAAGAAATAGAGAAATCGCTGAATCGCC 270  
 DB 21 TyrTyrAspProThrLysLysLysLysSerAspGluAspPheGluCysGly 40  
 QY 271 AGGTATCGCAAGCTCTTCAACACGACGACCATGGCATTTGTGTACTATAAAG 330  
 DB 41 ArgLeuSerProSerSerValGlySerGluProThrLysPheLeuValIleGlnAsnLys 60  
 QY 331 GATTAAAAAACAATTCAGTCAGTCAGCTACCTTTAATGAAGAAATGATTAAGCGCT 390  
 DB 61 ThrLeuArgGluLysMetLysProPheSerTrpLysMetLeuGlnLeuAspAsnLys 80  
 QY 391 TCAGGTATGATGGTGTATGCTCTTTAAGACCTAGCGATTTGTACCACAGCGCCATTAC 450  
 DB 81 SerHisLeuValValIleLeuAlaLysLys-----AsnAlaArgTyr 94  
 QY 451 ATGCAAAACCTTTACCGGAGTCCTTATAGGTAGATGATCCCTCTTTGCTCAATG 510  
 DB 95 AspSerProPhePheValAspValMetAlaArgLysGlyLeuAsnAlaGluGlnGln 114  
 QY 511 CTGGCGTGAATTCACACGACGATGCAAGAGATAGAAAGCTATATTTTAGAG----- 564  
 DB 115 AlaAlaLeuThrLysTyrLysAlaLeuGlnGluAspMetLysLeuLeuGluAsnAsp 134  
 QY 565 -----CAATGCTATATCGCTGTGGGCGCAAAATTGCGATG 597  
 DB 135 ArgThrLeuPheAspTrpCysSerLysGlnThrTyrIleAlaLeuAlaAsnMetLeuThr 154

QY 598 GCGCGTGAATTAATGGATTGGATAGTATGTCATTATTCGAGCGTTTGATCCTTTAAAGTG 657  
 DB 155 GlyAlaSerAlaLeuGlyIleAspSerCysProIleGluGlyPheHisTyrAspLysMet 174  
 QY 658 GGTGAAGTTTAA---GAAGACGATATCAATAAGCCTAAA-----ATCGCATGCTTG 705  
 DB 175 AsnGluCysLeuAlaGluGluGlyLeuPheAspProGlnGlnTyrAlaValSerValAla 194  
 QY 706 ATCGCTTTGGCAAGAGGGTGGCAGACGCGCAAGCAAAATCAAGAAATCAAAAGTTGAT 765  
 DB 195 AlaThrPheGlyTyrArgSerArgAspIleAlaLysLysSerArgLysGlyLeuAspGlu 214  
 QY 766 GCGATTACTTGGTTG 780  
 DB 215 ValValLysTrpVal 219  
 RESULT 5  
 ID NFNB\_ENTCL STANDARD; PRT; 217 AA.  
 AC Q01234;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Oxygen-insensitive NAD(P)H nitroreductase (EC 1.-.-.-).  
 GN NFNB OR NFS1.  
 OS Enterobacter cloacae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Enterobacter.  
 OX NCBI\_TaxID=550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 43560 / 96-3;  
 RX MEDLINE=91154203; PubMed=1999405;  
 RA Bryant C., Hubbard L., McElroy W.D.;  
 RT "Cloning, nucleotide sequence, and expression of the nitroreductase  
 gene from Enterobacter cloacae.";  
 RL J. Biol. Chem. 266:4126-4130(1991).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=91154202; PubMed=1999405;  
 RA Bryant C., Deluca M.;  
 RT "Purification and characterization of an oxygen-insensitive NAD(P)H  
 nitroreductase from Enterobacter cloacae.";  
 RL J. Biol. Chem. 266:4119-4125(1991).  
 CC -1- FUNCTION: REDUCTION OF A VARIETY OF NITROAROMATIC COMPOUNDS USING  
 NADH (AND TO LESSER EXTENT NADPH) AS SOURCE OF REDUCING  
 EQUIVALENTS; TWO ELECTRONS ARE TRANSFERRED.  
 CC -1- COPACTOR: FMN.  
 CC -1- PATHWAY: THE NITROREDUCTASE MIGHT BE INVOLVED IN THE QUINONE  
 METABOLISM. IT SHOWS FUNCTIONAL SIMILARITY TO MAMMALIAN  
 QUINONE REDUCTASES.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SIMILARITY: Belongs to the nitroreductase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M63808; AA62801.1; -  
 DR FIR; A38686; A38686.  
 DR PDB; 1KQB; 03-APR-02.  
 DR PDB; 1KQC; 03-APR-02.  
 DR PDB; 1KOD; 03-APR-02.  
 DR PDB; 1NEC; 31-MAR-00.  
 DR InterPro: IPR000415; Nitroreductase.  
 DR Pfam; PF00881; Nitroreductase, 1.  
 DR Oxidoreductase; NAD; NADP; Flavoprotein; FMN; 3D-structure.  
 FT NP\_BIND 153 158 NAD OR NADP (BY SIMILARITY).  
 FT

SQ	SEQUENCE	217 AA; 23950 MW; C42AA3DB184D5D9B CRC64;	
RP	SEQUENCE FROM N.A.		
RC	STRAIN=LT2 / SGSC1412 / ATCC 700720;		
RX	MEDLINE=21534948; PubMed=11677609;		
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,		
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,		
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,		
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,		
RA	Waterston R., Wilson R.K.;		
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium		
RT	LT2.";		
RL	Nature 413:852-856 (2001).		
CC	-!- FUNCTION: REDUCTION OF A VARIETY OF NITROAROMATIC COMPOUNDS USING		
CC	NADH (AND TO LESSER EXTENT NADPH) AS SOURCE OF REDUCING		
CC	EQUIVALENTS; TWO ELECTRONS ARE TRANSFERRED. CAPABLE OF REDUCING		
CC	NITROFURAZONE (BY SIMILARITY).		
CC	-!- COFACTOR: FMN		
CC	-!- SUBUNIT: Monomer (By similarity).		
CC	-!- SIMILARITY: Belongs to the nitroreductase family.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC			
DR	EMBL; X17250; CAA35113.1; -.		
DR	EMBL; AE008722; AAL19529.1; -.		
DR	PIR; S08397; S08397.		
DR	HSP; P38489; IDS7.		
DR	StyGene; SG10246; nfnB.		
DR	InterPro; IPR000415; Nitroreductase.		
DR	Pfam; PF00881; Nitroreductase; 1.		
KW	Oxidoreductase; NAD; NADP; Flavoprotein; FMN; Complete proteome.		
FT	NP BIND 153 158 NAD OR NADP (BY SIMILARITY).		
SQ	SEQUENCE 217 AA; 23955 MW; E90E9E05A0826D0F CRC64;		
	Alignment Scores:		
	Pred. No.: 7,55e-11 Length: 217		
	Score: 196.50 Matches: 52		
	Percent Similarity: 48.35% Conservative: 36		
	Best Local Similarity: 28.57% Mismatches: 55		
	Query Match: 12.60% Indels: 39		
	DB: 1 Gaps: 5		
US-09-647-661-1 (1-900) x NFNB_SALTY (1-217)			
QY	193 GAGGCCATTCTTGCAGATGTTTGACAGCCATTATGAGTTTCTAGTGAAGAATTAGAA	252	
Db	9 GlnArgTyrSerThrLysAlaPheAspProSerLysLysLeuThrAlaGluAlaAsp	28	
QY	253 GAAATCGCTGAAATCGCCAGGCTATCGCAAGCTCTTACACACGACGCCATGCGATTTT	312	
Db	29 LysileLysThrLeuLeuGlnTyr-SerProSerSerThrAsnSerGlnProTrpHisPhe	48	
QY	313 GTGATGTTACTAATAGCAATTTAAACAAATTCGAGTGCACAGC-----	360	
Db	49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnTyr	68	
QY	361 TACTTTAATGAAGAAATGATTAAAGCGCTTCACCGCTTAATGCGTATGCTCTTTAAGA	420	
Db	69 ThrPheAsnGluArgLysMetLeuAspAlaSerHisValValPheCysAla-----	86	
QY	421 CCTAGCGAGTTGTTACACACAGCGCATTTACATGCAAAACCTTTACCGGAGCTTTATAAG	480	
Db	87 -----LysThrAlaMetAspAspAlaTrpLeu	95	
QY	481 GTTAGAGTGATC-----	501	
Db	96 GluArgValValAspGlnGluAspAlaAspGlyArgPheAlaThrProGluAlaLysAla	115	
QY	502 GCTCAATGCTTGGCGTGAGATTCAACACAGCATGCAAAAGATTA-----	546	
Db	116 AlaAsnHisLysGlyArgThrTyrPheAlaAspMetHisArgValAspLeuLysAspAsp	135	
QY	547 GAAAGCTATATTGAGCAATGCTATATCGTGGGCAAAATTTGATGGCGGTGAGC	606	
Db	136 AspGlnTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeuGlyValGly	155	
QY	607 TTAATGGATGGATGCTGATGCTATTGAGGCTTTGATCCTTTAAAGTGGGTGAATT	666	
Db	156 AlaMetGlyLeuAspAlaValProLleGluGlyPheAsp-----AlaAlaIle	171	
QY	667 TTAGAAGAG 675		
Db	172 LeuAspGlu 174		
RESULT 6			
NFNB_SALTY			
ID	NFNB_SALTY STANDARD; PRT; 217 AA.		
AC	P1588;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Oxygen-insensitive NAD(P)H nitroreductase (EC 1.-.-.-).		
GN	NFNB OR STW0578.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmonella.		
NCBI_TaxID=602;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=TA1538;		
RX	MEDLINE=90192100; PubMed=2179862;		
RA	Watanabe M., Ishidate M., Noimi T.;		
RT	"Nucleotide sequence of Salmonella typhimurium nitroreductase gene.";		
RL	Nucleic Acids Res. 18:1059-1059 (1990).		
RN	[2]		

Db 116 AlaAsnAspLysGlyArgGlyPheAlaAspMethHisArgValSerLeuLysAspAsp 135  
 Qy 547 GAAAGCTATATTGAGCAATGCTATATCGCTGTGGGCAAAATTTGCATGGCGTGAGC 606  
 Db 136 HisGlnTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeuGlyValAla 155  
 Qy 607 TTAATGGATGGATGCTGATCATATTGGAGGCTTTCATCTTTAAAGTGGGTGAAGTT 666  
 Db 156 AlaMetGlyLeuAspAlaValProIleGluGlyPheAsp-----AlaGluVal 171  
 Qy 667 TTAGAA 672  
 Db 172 LeuAsp 173

RESULT 7  
 NFNB\_ECOLI STANDARD; PRT; 217 AA.  
 ID NFNB\_ECOLI  
 AC P38489; P19575;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Oxygen-insensitive NAD(P)H nitroreductase (EC 1.-.-.-) (FMN-dependent  
 DE nitroreductase) (Dihydropteridine reductase) (EC 1.6.99.7).  
 GN NFNB OR NFBI OR NFSE OR NTR OR DPRA OR B0578.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / C500;  
 RA Zenko S., Koike H., Tanokura M., Saigo K.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B;  
 RX Michael N.P., Brehm J.K., Anlezark G.M., Minton N.P.;  
 RA "Physical characterisation of the Escherichia coli B gene encoding  
 RT nitroreductase and its over-expression in Escherichia coli K12.";  
 RL FEMS Microbiol. Lett. 124:195-202(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sempel G., Seki Y., Takami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horuchi T.;  
 RP "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [6]  
 RP CHARACTERIZATION, AND SEQUENCE OF 1-31 AND 139-180.

RC STRAIN=B;  
 RX MEDLINE=93111206; PubMed=1472094;  
 RA Anlezark G.M., Melton R.G., Sherwood R.F., Coles B., Friedlos F.,  
 RA Knox R.J.;  
 RT "The bioactivation of 5-(aziridin-1-yl)-2,4-dinitrobenzamide  
 RT (CB1954) -- I. Purification and properties of a nitroreductase enzyme  
 RT from Escherichia coli -- a potential enzyme for antibody-directed  
 RT enzyme prodrug therapy (ADPPT).";  
 RL Biochem. Pharmacol. 44:2289-2295(1992).  
 RN [7]  
 RP SEQUENCE OF 1-12.  
 RC STRAIN=K12 / EMG2;  
 RX Link A.J., Robison K., Church G.M.;  
 RA "Comparing the predicted and observed properties of proteins encoded  
 RT in the genome of Escherichia coli K-12.";  
 RL Electrophoresis 18:1259-1313(1997).  
 RN [8]  
 RP SEQUENCE OF 1-20, AND CHARACTERIZATION AS A DIHYDROPTERIDINE REDUCT.  
 RX MEDLINE=89076221; PubMed=3060113;  
 RA Vasudevan S.G., Shaw D.C., Armarego W.L.F.;  
 RT "Dihydropteridine reductase from Escherichia coli.";  
 RL Biochem. J. 255:581-588(1988).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.06 ANGSTROMS).  
 RC STRAIN=B;  
 RX MEDLINE=20476666; PubMed=11020276;  
 RA Parkinson G.N., Skelly J.V., Neidle S.;  
 RT "Crystal structure of FMN-dependent nitroreductase from Escherichia  
 RT coli B: a prodrug-activating enzyme.";  
 RL J. Med. Chem. 43:3624-3631(2000).  
 CC -1- FUNCTION: REDUCTION OF A VARIETY OF NITROAROMATIC COMPOUNDS USING  
 CC NADH (AND TO LESSER EXTENT NADPH) AS SOURCE OF REDUCING  
 CC EQUIVALENTS; TWO ELECTRONS ARE TRANSFERRED. CAPABLE OF  
 CC NITROFURAZONE, QUINONES AND THE ANTI-TUMOR AGENT CB1954 (5-  
 CC (AZIRIDIN-1-YL)-2,4-DINITROBENZAMIDE). THE REDUCTION OF CB1954  
 CC RESULTS IN THE GENERATION OF CYTOTOXIC SPECIES.  
 CC -1- CATALYTIC ACTIVITY: NAD(P)H + 6,7-dihydropteridine = NAD(P) (+) +  
 CC 5,6,7,8-tetrahydropteridine.  
 CC -1- COFACTOR: FMN.  
 CC -1- SUBUNIT: Monomer or homodimer.  
 CC -1- SIMILARITY: Belongs to the nitroreductase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D25414; BAA05004.1; -.  
 CC EMBL; J07860; AAC43263.1; -.  
 CC EMBL; AE000163; AAC73679.1; -.  
 CC EMBL; U82598; AAB40776.1; -.  
 CC EMBL; D90700; BAA35218.1; -.  
 CC EMBL; A23284; CAA01666.1; -.  
 CC EIR; I67685; I67685.  
 CC EIR; S01818; S01818.  
 CC PDB; 1DS7; 12-JUL-00.  
 CC PDB; 1ICR; 23-MAY-01.  
 CC PDB; 1ICU; 23-MAY-01.  
 CC PDB; 1ICV; 23-MAY-01.  
 CC SWISS-2DPAGE; P38489; COLI.  
 CC EcoGene; EG20151; nfnB.  
 CC InterPro; IPR000415; Nitroreductase.  
 CC Pfam; PF00881; Nitroreductase; 1.  
 KW Oxidoreductase; Nitroreductase; FMN; Complete proteome;  
 3D-structure.  
 FT NP\_BIND 153 158 NAD OR NADP (BY SIMILARITY).  
 FT S 5 S -> C (IN REF. 8).  
 FT CONFLICT 10 12 RHS -> CIV (IN REF. 8).  
 FT CONFLICT 19 19 S -> M (IN REF. 8).  
 FT

```

FT FT CONFLICT 21 21 MISSING (IN REF. 6).
FT FT CONFLICT 28 28 E -> D (IN REF. 6).
FT FT CONFLICT 180 180 E -> I (IN REF. 6).
FT FT HELIX 3 9
FT FT STRAND 12 12
FT FT TURN 16 16
FT FT TURN 18 19
FT FT TURN 24 35
FT FT TURN 36 36
FT FT TURN 40 42
FT FT STRAND 46 51
FT FT HELIX 54 61
FT FT TURN 62 63
FT FT HELIX 66 71
FT FT HELIX 72 77
FT FT STRAND 80 87
FT FT HELIX 92 104
FT FT TURN 105 106
FT FT TURN 111 129
FT FT TURN 130 132
FT FT TURN 135 157
FT FT TURN 158 158
FT FT STRAND 160 161
FT FT STRAND 164 164
FT FT TURN 169 175
FT FT TURN 176 177
FT FT HELIX 179 181
FT FT TURN 182 182
FT FT STRAND 183 192
FT FT STRAND 194 194
FT FT TURN 196 197
FT FT HELIX 199 201
FT FT TURN 202 202
FT FT TURN 210 213
FT FT STRAND 214 217
SQ SEQUENCE 217 AA; 23905 MW; A516CEFC3D46AEAC CRC64;

Alignment Scores:
Pred. No.: 4.37e-10
Score: 188.50
Percent Similarity: 47.73%
Best Local Similarity: 27.84%
Query Match: 12.09%
DB: 1

US-09-647-661-1 (1-900) x NFNB_ECOLI (1-217)
Qy 193 GAGCGCCATCTTCACAGATGTTTGACGACCATATAGAGTTTCTAGTGAAGAATTAGAA 252
Db 9 LysArgHisSerThrLysAlaPheAspAlaSerLysLysLeuThrProGluGlnAlaGlu 28
Qy 253 GAAATCGCTGAAATCGCCAGCGCTATCGCAAGCTCTTACACAGCGACGCGCATGCGATTTT 312
Db 29 GlnIleLysThrLeuLeuGlnTyrSerProSerSerThrAsnSerGlnProTrpHisPhe 48
Qy 313 GTGATGTTACTAATAAGGATTTTAAACAAATTCAGTCGACACGAC----- 360
Db 49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnTyr 68
Qy 361 TACTTTTATGAGAAATGATTAAGCGCTTCACGCTTATGTTGTTGTTGTTGTTTAAAGA 420
Db 69 ValPheAsnGluArgLysMetLeuAspAlaSerHisValValPheCysAlaLysThr 88
Qy 421 CCTAGCGAGTTGTTTACCACACGGCCATTACATGCAAAACCTTTAC----- 465
Db 89 Ala-----MetAspAspValTrpLeuLysLeuValVal 99
Qy 466 -----CCGAGTCTCTATAGGTTAGTAGTATC 492
Db 100 AspGlnGluAspAlaAspGlyArgPheAlaThrProGluAla----- 113
Qy 493 CCTTCTTTTGTCAAATGCTTGGCGTGGAGATTCAACACGACGATCGCAAGATTAGAA--- 549

```

```

Db 114 ---LysAlaAlaAsnAspLysGlyArgLysPhePheAlaAspMetHisArgLysAspLeu 132
Qy 550 -----AGCTATATTTTACGCAATGCTATATCGTGTGGCGCAAAATTTGCAATG 597
Db 133 HisAspAspAlaGluTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeu 152
Qy 598 GCGGTGAGCTTAATGGGATTTGATAGTTGCTATTATTGGAGGCTTTGAT 645
Db 153 GlyValAlaAlaLeuGlyLeuAspAlaValProIleGluGlyPheAsp 168

RESULT 8
V12C_BACSU
ID V12C_BACSU STANDARD; PRT; 201 AA.
AC P81102;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative NAD(P)H nitroreductase 12C (EC 1.1.1.1) (Vegetative protein
DE 12C) (VEG12C).
OS YODC OR BSU19550.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Wambutt R., Wedler H., Lapidus A., Sorokin A., Ehrlich S.D.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the odbAB and espC loci cloned in a yeast artificial chromosome.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunat F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.X., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta E., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrito P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takenaru K.,
RA Tasato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viarri A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 1-23.
RC STRAIN=168 / JH642;
RA Graumann P.L., Schmid R., Marahiel M.A.;
RL Submitted (OCT-1997) to Swiss-Prot.
RN [4]
RP CHARACTERIZATION.
RC STRAIN=168 / JH642;
RX MEDLINE=96345629; PubMed=8755892;

```

175 MetLeuIserIecIyIysaIaValIysProAlaHisGlnSerAsnArgLeuProLeu 194

760 GTTGATCGCATTAACCTGGTGG 780

195 SerIysValSerThrIrpLeu 201

RESULT 9

YDGI\_BACSU STANDARD; PRT; 209 AA.

AC P96707;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Putative NAD(P)H nitroreductase ydgi (EC 1.-.-.-).

GN YDGI OR BSU05660.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]\_TaxID=1423;

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Kasahara Y., Nakai S., Lee S., Sadaie Y., Ogasawara N.;

RT "A 148 kbp sequence of the region between 35 and 47 degree of the

RL Bacillus subtilis genome.";

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,

RA Azevedo V., Berta M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Bruchsi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,

RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,

RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerf-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetalle D., Porwollik S., Prescott A.M.,

RA Presecan E., Fujic P., Furnelle B., Rpoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serrori P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,

RA Winkers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zunststein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus

RT subtilis.";

RL Nature 390:249-256 (1997).

CC -!- COFACTOR: FMN (Potential).

CC -!- SIMILARITY: Belongs to the nitroreductase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/commence/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AB001488; BAA19399.1; -.

DR EMBL; Z99107; CAB21385.1; -.

DR PIR: C69783; C69783.  
 DR HSPF; Q60049; INOX.  
 DR Subtilist; BG12176; ydgi.  
 DR InterPro; IPR000415; Nitroreductase.  
 DR Pfam; PF00881; Nitroreductase; 1.  
 KW Hypothetical protein; Oxidoreductase; NAD; NADP; Flavoprotein; FMN;  
 KW Complete proteome.  
 SQ SEQUENCE 209 AA; 23359 MW; A40E9D459F560597 CRC64;

Alignment Scores:  
 Pred. No.: 1,47e-08 Length: 209  
 Score: 172.50 Matches: 52  
 Percent Similarity: 49.07% Conservatives: 53  
 Best Local Similarity: 24.30% Mismatches: 80  
 Query Match: 11.06% Indels: 29  
 DB: 1 Gaps: 9

US-09-647-661-1 (1-900) x YDGI\_BACSU (1-209)

Qy 181 CAACATATAACAGCGCCCATCTTCGCAAGATGTTTACAGCCATTATGAGTTTCTAGT 240  
 Db 9 GluileMetLysGlyArgSerIleArgAsnTyrAspProAlaValLysIleSerLys 28

Qy 241 GAAGAATTAGAAATCGCTGAATCGCAGCGCTATCGCAAGCTCTTACACACGCGAG 300  
 Db 29 GluGluMetThrGluileLeuGluGluAlaThrThrAlaProSerSerValAlaGln 48

Qy 301 CCATGGCATTTTGTGATGTTTACTAATAAGGATTTAAACAAATGCGAGTCACAGC 360  
 Db 49 ProTprArgPheLeuValIleAspSerProGluGlyLysGluLysLeuAlaProLeuAla 68

Qy 361 TACTTTAATGAAGAAATGATTAAAGCGCTTCAGCGTTAATGCTGCTCTTTAAGA 420  
 Db 69 SerPheAsnGlnThrGlnValThrThrSerAlaValIleAlaValPheAla----- 86

Qy 421 CCTAGCGAGTTGTTACACACGCGCATATACATGCAAAACCTTTAC----- 465  
 Db 87 -----AspMetAsnAsnAlaAspTyrLeuGluGluIleTyrSerLysAlaValGlu 103

Qy 466 -----CGGAGTCTTATPAGGTTAGAGTATCCCTCTCTTTTGT----- 504  
 Db 104 LeuGlyTyrMetProGlnGluValLysAspArgGlnIleAlaLeuThrAlaHisPhe 123

Qy 505 CAATGCTTGGCGTGAGATTCAACACACAGCATGCAAGATAGAACTATATTTAGAG 564  
 Db 124 GluLysLeuProAlaGlnValAsnArg-----GluThrIleLeuileAsp 138

Qy 565 CAATGCTATATCGCTGTGGGCAATTTGTCATGGCGTGAGCTTAATGGGATTGGAGT 624  
 Db 139 GlyGlyLeuValSerMet---GlnLeuMetLeuThrAlaArgAlaHisGlyTyrAspThr 157

Qy 625 TGCATTATTGAGCGCTTGATCCCTTTAAAGTGGGTGAAGTT-----TTAGAGAGCGT 678  
 Db 158 AsnProIleGlyGlyTyrAspLysGluAsnIleAlaGluThrPheGlyLeuAspLysGlu 177

Qy 679 ATCAATAAGCCATAAATCGCATGCTTGCATCGCTTTGGCAAGGGTGGCAGAGCGAGC 738  
 Db 178 ArgTyrValPro-----ValMetLeuLeuSerIleGlyLys---AlaAlaAspGluGly 194

Qy 739 CAAAAATCAAGAAATCAAAAGTTGATCGGATTACT---TGG 777  
 Db 195 TyrAlaSerTyrArgLeuProIleAspThrIleAlaGluTrp 208

RESULT 10  
 DRGA\_SINY3  
 ID DRGA\_SINY3 STANDARD; PRT; 210 AA.  
 AC Q55233; P73850;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein drga.  
 GN DRGA OR SLE1719.  
 OS Synechocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98316704; PubMed=9654132;  
 RA Blanskaya I.V., Chesnavichene E.A., Vernotte C., Astier C.;  
 RT "Resistance to nitrophenolic herbicides and metronidazole in the  
 cyanobacterium Synechocystis sp. PCC 6803 as a result of the  
 inactivation of a nitroreductase-like protein encoded by drga gene.";  
 RL FEBS Lett. 428:188-192(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 Miyajima N., Hikosawa M., Sugita M., Sasamoto S., Kimura T.,  
 Hosouchi I., Matsuno A., Muraki A., Nakazaki N., Nardo K.,  
 Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
 Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 RN [3]  
 RP SEQUENCE OF 1-17.  
 RX MEDLINE=97443974; PubMed=9298645;  
 RA Sazuka T., Ohara O.;  
 RT "Towards a proteome project of cyanobacterium Synechocystis sp.  
 strain PCC6803: linking 130 protein spots with their respective  
 genes";  
 RL Electrophoresis 18:1252-1258(1997).  
 CC -!- FUNCTION: CONTROLS RESISTANCE TO THE HERBICIDE DINOSEB AND  
 METRONIDAZOLE. INVOLVED IN DETOXIFICATION OF DINOSEB VIA THE  
 REDUCTION OF THE NITRO GROUP(S) AND THIS PROCESS IS ACCOMPANIED BY  
 THE FORMATION OF TOXIC SUPEROXIDE ANIONS.  
 CC -!- COFACTOR: FMN (Potential).  
 CC -!- SIMILARITY: Belongs to the nitroreductase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL; J29426; AAA91952.1; -.  
 DR EMBL; D90910; BAA17909.1; -.  
 DR InterPro; IPR000415; Nitroreductase; 1.  
 DR Pfam; PF00881; Nitroreductase; 1.  
 KW Herbicide resistance; Oxidoreductase; NAD; NADP; Flavoprotein; FMN;  
 KW Complete proteome.  
 FT CONFLICT 31 31 H -> L (IN REF. 2).  
 FT CONFLICT 45 45 L -> H (IN REF. 2).  
 FT CONFLICT 183 210 FPKRRNSFGRIPLGKLLCLTKVWCLAI -> RAKGGQTPL  
 FT BELVWENSF (IN REF. 2).  
 SQ SEQUENCE 210 AA; 23703 MW; C29A1293996EBB95 CRC64;

Alignment Scores:  
 Pred. No.: 2,83e-08 Length: 210  
 Score: 169.50 Matches: 53  
 Percent Similarity: 46.11% Conservatives: 36  
 Best Local Similarity: 27.46% Mismatches: 87  
 Query Match: 10.87% Indels: 17  
 DB: 1 Gaps: 5

US-09-647-661-1 (1-900) x DRGA\_SINY3 (1-210)

Qy 193 GAGCGCCCATCTTCGCAAGATGTTTACAGCCATTATGAGTTTCTAGTGAAGATTAGAA 252  
 Db 9 GlnArgSerValLysHisPheAspProAspHisArgLeuThrAlaGluGluArg 28

Qy 253 GAAATCGCTGAATCGCAGGCTATCGCAGCTCTTACACACGCGCATTTT 312  
 Db 253 GAAATCGCTGAATCGCAGGCTATCGCAGCTCTTACACACGCGCATTTT 312





DB:	1	Gaps:	7
US-09-647-661-1	(1-900) x FRA1_VIBF1	(1-217)	
QY	187	TTAAACGAGCGCATCTTTCGCAAGATGTTTGACAGCCATTAGTTCCTAGTGAAGAA	246
Db	8	LeuGluAsnArgTyrThrSerLysLysTyrAspProSerLysLysValSerGlnGluAsp	27
QY	247	TTAGAGAAATCGCTGAATCGCCAGGTATGTCGCCAGCTTTCAACACGACGCCATGG	306
Db	28	LeuAlaValLeuGluAlaLeuArgLeuSerAlaSerSerIleAenSerGlnProTyr	47
QY	307	CATTTCGTGATCGTTACTAATAAGCATTTAAAAACAAATTCGAGTCACAGCTAC	363
Db	48	LysPheIleValle---GluSerAspAlaLysGlnArgMetHisAspSerPheAla	66
QY	364	-----TTTAATGAAGAAATGATTTAAAGCGCTTCAGCGTTAAGTGGTATGC	411
Db	67	AsnMetHisGlnPheAsnGlnProHisIleLysAlaCysSerHisValIleLeuPheAla	86
QY	412	TCCTTAAGACCTAGCGAGTGTTCACACGCGCATTCATCATGCAAAACCTTTAC	465
Db	87	AsnLysLeu-----SerTyrThrArgAspAspTyrAspVal	98
QY	466	-----CCGGAGTCTTTATAGGTTAGAGTGATC	492
Db	99	ValLeuSerLysAlaValAlaAspLysArgIleThrGluGlnLysGluAlaAlaPhe	118
QY	493	CCTCTCTTTGCTCAAAATGTTGGGTGAGATTCACACAGCATGCCAAGATTAGAAGC	552
Db	119	AlaSerPhe---LysPheValGluLeuAsnLysCysAspGluAsnGlyGluHis---	136
QY	553	TATATTTTAGCAATGCTATATCGTGTGGGGCAAAATTCGATGGCGTGGCTTAATG	612
Db	137	TyrThrLysProGlnAlaTyrLeuAlaLeuGlyAsnAlaLeuHisThrLeuAlaArgLeu	156
QY	613	GGATTGGATGTGCATTTATGAGCGTTTGTATCTTTAAAGTGGGTGAAGTTTGA	672
Db	157	AsnIleAspSerThrThrMetGluGlyIleAspProGluLeuLeuSerGluIlePheAla	176
QY	673	GAGCGTATCAATAAGCCTAAATCGCATCGTTCGCTTGGGC-----	717
Db	177	AspGluLeuLysGlyTyrGluCysHisValAlaLeuAlaIleGlyTyrHisHisProSer	196
QY	718	AAGAGGGTGGCAGACGCGCAAAATCAAGAAATCAAAAGTTGATCGCATCTTGG	777
Db	197	GluAspTyrAsnAlaSerLeuProLysSerArgLysAlaPheGluAspValIleThrIle	216
QY	778	TTG 780	
Db	217	Leu 217	
RESULT 12			
YMF7_ARCFU			
ID_YMF7_ARCFU	STANDARD;	PRT;	174 AA.
AC	O28017;		
DT	15-DEC-1998	(Rel. 37, Created)	
DT	15-DEC-1998	(Rel. 37, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last annotation update)	
DE	Putative NADH dehydrogenase/NAD(P)H nitroreductase AF2267		
DE	(EC 1.-.-.-)		
GN	AF2267.		
OS	Archaeoglobus fulgidus.		
OC	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;		
OC	Archaeoglobaceae; Archaeoglobus.		
OC	NCBI_TaxID=2234;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VC-16 / DSM 4304 / ATCC 49558;		
RA	MEDLINE=98049343; PubMed=9389475;		
RA	Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,		
RA	Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,		
RA	Richardson D.L., Karlavaage A.R., Graham D.E., Kyrpides N.C.,		

Db 104 IleLeuLeuAlaAlaArgCysLeuGlyIleGlySerValTrpLeuGlyValTyProArg 123  
QY 649 TTAAGAGTGGTGAAGCTTTTAAAGAGCGGTATCAATAGCTATAA-----ATCGCA 699  
Db 124 GluGluArgMetLysAlaLeuArgGluLeuGlyLeuProGluAenIleValValPhe 143  
QY 700 TGCTTGATCGCTTTGGGCG-----AAGAGGGTGGCAGAGGAGCGCAAA 744  
Db 144 SerValValSerLeuGlyTyProLysAspGluLysAspPheTyGluAlaAspArg 163  
QY 745 TCAGAGAAATCAAAAGTTGATCGCATTTCTGG 777  
Db 164 PheAsnProAspArgIleHisArgGluLysTrp 174  
RESULT 13  
YDFN\_BACSU  
ID YDFN\_BACSU STANDARD; PRT; 206 AA.  
AC P96692;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Putative NAD(P)H nitroreductase ydfN (EC 1.-.-.-).  
GN YDFN OR BSU05480.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RT "A 148 kbp sequence of the region between 35 and 47 degree of the  
RT Bacillus subtilis genome."  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunet F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borris L., Bourrier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Duxterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppi G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kashara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,  
RA Presecan E., Fujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tononni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.P., Zumbstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis".  
RL Nature 390:249-256(1997).  
CC -I- COFACTOR: FMN (Potential).  
CC -I- SIMILARITY: Belongs to the nitroreductase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC  
DR EMBL; AB001486; BAA1382.1; ALT\_INIT.  
DR EMBL; Z99106; CAB12355.1; --.  
DR PIR; D69781; D69781.  
DR Subtilist; BG12161; ydfN  
DR InterPro; IPR000415; Nitroreductase.  
DR Pfam; PF00881; Nitroreductase; 1.  
DR KW Hypothetical protein; Oxidoreductase; NAD; NADP; Flavoprotein; FMN;  
KW Complete proteome.  
SQ SEQUENCE 206 AA; 23559 MW; 99BFE7FDBD9075FC CRC64;  
Alignment Scores:  
Pred. No.: 3.19e-05 Length: 206  
Score: 137.50 Matches: 51  
Percent Similarity: 44.88% Conservative: 41  
Best Local Similarity: 24.88% Mismatches: 70  
Query Match: 8.82% Indels: 43  
DB: 9 Gaps: 9  
US-09-647-661-1 (1-900) x YDFN\_BACSU (1-206)  
QY 184 CTATTAAACGAGCGCCATCTTGGCAAGATGTTGACAGCATATGAGTTTCTAGTGA 243  
Db 7 LeuValAsnGluArgSerAlaSerAsnPheLeuSerGlyHisProIleThrLysGlu 26  
QY 244 GAATTAGAAAGAAATCGCTGAAATGCCAGGCTATCCCAAGCTTCAACACGAGCCCA 303  
Db 27 AspLeuAsnGluMetPheGluLeuValAlaLeuAlaProSerAlaPheAsnLeuGlnHis 46  
QY 304 TGGCATTTGTGATGTTACTTAATAGGATTTAAA-----AAACAATTCAGTG 354  
Db 47 ThrLysTyValThrValLeuAspGlnAspValLysGluLysLeuLysGlnAlaAlaAsn 66  
QY 355 CACAGCTACTTTAATGAAGAAATGTTAAAGCGTTTCAGCGTTAATGCTGCTGCTCT 414  
Db 67 GlyGlnTy-----LysValValSerSerSerAlaValLeuLeuValLeuGly 82  
QY 415 TTAAGACCTAGCGAGTTGTACACACGCGCCATTACATGCAA-----AACCTTACCG 468  
Db 83 AspLysGlnAla-----TyrGlnGlnAlaAlaAspIleTyr----- 94  
QY 469 GAGCTTTATAGGTTAGAGTATCCCTTTCTTTCCTCAATGCTTGGCGTG----- 519  
Db 95 GluGlyLeuLysVal-----LeuGlyIleLeuAsnLys 105  
QY 520 ---AGATTCAACACACGACGATGCAAGATTAGAAAGCTAT----- 555  
Db 106 GlnGluTyAspHisMetValGlnAspThrValSerPheTyGluAsnArgGlyGluGln 125  
QY 556 -----ATTATTAGAGCATGCTATATCGCTGCTGGGCAATTTGCATGGCG 600  
Db 126 PheLysArgAspGluAlaIleAlaSerLeuSerAlaMetMetMetMetLeuSer 145  
QY 601 GTGAGCTTAATGGGATTTGATGTTGATTTGAGGCTTTGATCCTTTAAAGTGGGT 660  
Db 146 AlaLysGluLysGlyTrpAspThrCysProMetIleGlyPheAspAlaGluAlaValLys 165  
QY 661 GAAGTTTAAAGAGCGGTATCAATAGCTTAAATCGCATGCTTGTGCTTGGCGCAG 720  
Db 166 ArgIleLeuAsnIleAsp---AspGlnPheGluValMetMetMetIleThrIleGlyLys 184  
QY 721 AGGCTGGCAGGAGCG 735  
Db 185 GluLysThrGluSer 189  
RESULT 14  
NOX\_THETH  
ID \_NOX\_THETH STANDARD; PRT; 205 AA.

AC Q60049; Q53306;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE NADH dehydrogenase (EC 1.6.99.3) (H(2)O(2) forming NADH oxidase)  
 GN (NADH:oxygen oxidoreductase).  
 GN NOX.  
 OS Thermus thermophilus.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 OC Thermus.  
 OX NCBI\_TaxID=274;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=HB8 / ATCC 27634;  
 RC MEDLINE=92249333; PubMed=1577004;  
 RA Park H.-J., Kreutzer R., Reiser C.O.A., Sprinzl M.;  
 RT "Molecular cloning and nucleotide sequence of the gene encoding a  
 RT H2O2-forming NADH oxidase from the extreme thermophilic Thermus  
 RT thermophilus HB8 and its expression in Escherichia coli.";  
 RL Eur. J. Biochem. 205:875-879(1992).  
 RN [2]  
 RN ERRATUM.  
 RX MEDLINE=93170329; PubMed=8436145;  
 RA Park H.-J., Kreutzer R., Reiser C.O.A., Sprinzl M.;  
 RL Eur. J. Biochem. 211:909-909(1993).  
 RN [3]  
 RX X-RAY CRYSTALLOGRAPHY (1.59 ANGSTROMS).  
 RC MEDLINE=96110872; PubMed=8846223;  
 RA Hecht H.J., Erdmann H., Park H.-J., Sprinzl M., Schmid R.D.;  
 RT "Crystal structure of NADH oxidase from Thermus thermophilus.";  
 RL Nat. Struct. Biol. 2:1109-1114(1995).  
 RN [4]  
 RN SEQUENCE OF 1-32, AND CHARACTERIZATION.  
 RC STRAIN=HB8 / ATCC 27634;  
 RC MEDLINE=92249332; PubMed=1577005;  
 RA Park H.-J., Reiser C.O.A., Kondrweit S., Erdmann H., Schmid R.D.,  
 RA Sprinzl M.;  
 RT "Purification and characterization of a NADH oxidase from the  
 RT thermophile Thermus thermophilus HB8.";  
 RL Eur. J. Biochem. 205:891-895(1992).  
 CC -!- FUNCTION: Thermotable enzyme that exhibits highest activity at pH  
 CC 5.0 and can oxidize either NADH or NADPH with a preference for  
 CC NADH. Can catalyze electron transfer from NADH to various electron  
 CC acceptors which include, in addition to molecular oxygen,  
 CC cytochrome c, 2,6-dichlorophenolindophenol, methylene blue,  
 CC ferricyanide or P-nitroblue tetrazolium.  
 CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.  
 CC -!- COFACTOR: Binds 1 FMN per subunit.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SIMILARITY: Belongs to the nitroreductase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 DR EMBL; X60110; CAA42707.1; -;  
 DR EMBL; S55441; AAB25458.1; -;  
 DR PDB; 1NOX; 12-MAR-97.  
 DR InterPro; IPR000415; Nitroreductase.  
 DR Pfam; PF00881; Nitroreductase; 1.  
 KW Oxidoreductase; NAD; Flavoprotein; FMN; 3D-structure.  
 FT HELIX 10 16  
 FT STRAND 23 23  
 FT HELIX 30 40  
 FT TURN 41 42  
 FT HELIX 46 48  
 FT STRAND 52 57  
 FT HELIX 60 69  
 FT TURN 70 72

FT TURN	74	74	74
FT HELIX	75	78	78
FT STRAND	81	87	87
FT HELIX	89	94	94
FT TURN	95	95	95
FT HELIX	96	98	98
FT TURN	99	99	99
FT TURN	102	103	103
FT TURN	106	106	106
FT HELIX	107	121	121
FT TURN	122	122	122
FT HELIX	125	149	149
FT TURN	150	151	151
FT STRAND	153	157	157
FT HELIX	162	169	169
FT TURN	170	170	170
FT TURN	173	174	174
FT STRAND	176	184	184
FT STRAND	186	186	186
FT HELIX	198	201	201
SQ SEQUENCE	205 AA;	22749 MW;	EE6CA2DFA85B50FE CRC64;

  

Alignment Scores:	3,97e-05	Length:	205
Pred. No.:	Score:	Matches:	46
Percent Similarity:	44.57%	Conservative:	36
Best Local Similarity:	25.00%	Mismatches:	83
Query Match:	8.76%	Indels:	19
DB:	1	Gaps:	5

  

US-09-647-661-1 (1-900) x NOX\_THETH (1-205)

QY	247	TTAGAAGAAATCGTGAATCCCGAGCTATCGCAAGCTCTTACACACGAGCGCATGG	306
Db	33	LeuArgGluileLeuGluAlaLeuArgAlaProSerAlaTrpAsnLeuGlnProTrp	52
QY	307	CATTITGATGCTTACTAATAAGGATTTAAAAAACAATTCAGTCGACAGCTACTTT	366
Db	53	ArgIleValValValArgAspProAlaThrLysArgAlaLeuArg---GluAlaAlaPhe	71
QY	367	AATGAAGAAATGATTAAAGCGCTTCAGGTTAATGGTGTATGCTCTTTAAGACCTAGC	426
Db	72	GlyGlnAlaHisValGluAlaProValValLeuValLeuTyAlaAspLeuGluAsp	91
QY	427	GAGTTGTTACCACGCGCATTCATGCAAAACCTTTACCGG-----	468
Db	92	AlaLeu-----AlaHisLeuAspGluValIleHisProGlyValGlnGlyGluArg	108
QY	469	---GAGTCTTATAAGGTTAGAGTGATCCCTCTTTTGTCTCAATGCTGGCGTGATTC	525
Db	109	ArgGluAlaGlnLysGlnAlaIleGlnArgAlaPheAlaAlaMet-----	123
QY	526	ACCACAGCATCAAGATTAGAAGCTATATTTAGACCATGCTATATCGCTGTGGG	585
Db	124	-----GlyGlnGluAlaArgLysAlaTrpAlaSerGlyGlnSerTyrlleLeuGly	141
QY	586	CAAAATTGTCATGGCGTGAGCTTAATGGGATTTAGTGTGATTTGTCATTTGAGGCTT	645
Db	142	TyrLeuLeuLeuLeuLeuGluAlaTyGlyLeuGlySerValProMetLeuGlyPheAsp	161
QY	646	CTTTTAAAGTGGTGAATTTTAGAGAGCGGTATCAATAGCCTAAATCGCATGCTTG	705
Db	162	ProGluArgValArgAlaIleLeu---GlyLeuProSerArgAlaAlaIleProAlaLeu	180
QY	706	ATCGCTTTGGCAAGAGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	765
Db	181	ValAlaLeuGlyTyPrProAlaGluGlyTyPrProSerHisArgLeuProLeuGluArg	200
QY	766	GGGATTACTTGG	777
Db	201	ValValLeuTrp	204

RESULT 15

```
QY 418 AGACCTAGGAGTTGTTACACACGGCCATTACATGCAAAACCTTTACCGGAGTCTTAT 477
DB 102 -----
QY 478 AAGGTTAGAGTATCCCTTCTTTTCTCAAAATGCTTGGCGTGAGATTCAACACACGATG 537
DB 103 -----
QY 538 CAAAGATTAGAAAGCTATATTTTATAGCAATGCTATATCGCTGTGGGCAAAATTTGCAATG 597
DB 107 GluLysIleProGlyPheTrpValGlnAspCysSerIleAlaSerGlnAsnIleLeuLeu 126
QY 598 GCGGTGAGCTTAATGGGATTGGATAGTTGCATTATTGGAGGCTTTGATCTCTTTAAAGTG 657
DB 127 AlaAlaHisSerLeuGlyLeuGlyAlaValTrpThrGlyValTyrPro----- 142
QY 658 GGTGAAGTTTATAGAGAGGATC 681
DB 143 -----LeuGluAspArgVal 147

Search completed: April 20, 2004, 19:25:32
Job time : 20 secs
```

```
Y120_METTH
ID Y120_METTH STANDARD; PRT; 191 AA.
AC O26223;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative NADH dehydrogenase/NAD(P)H nitroreductase (EC 1.-.-.-).
GN MTH120.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- COFACTOR: FMN (Potential).
CC -!- SIMILARITY: Belongs to the nitroreductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE000801; AAB84626.1; -.
DR PIR; B69027; B69027.
DR HSP; Q60849; INOX.
DR InterPro; IPR000415; Nitroreductase.
DR Pfam; PF00881; Nitroreductase; 1.
KW Hypothetical protein; Oxidoreductase; NAD; NADP; Flavoprotein; FMN;
KW Complete proteome.
FT NP_BIND 127 132 NAD OR NADP (BY SIMILARITY).
SQ SEQUENCE 191 AA; 21687 MW; FAE253F75DAD84BB CRC64;

Alignment Scores:
Pred. No.: Length: 191
Score: 126.00 Matches: 34
Percent Similarity: 41.07% Conservative: 25
Best Local Similarity: 20.24% Mismatches: 51
Query Match: 8.08% Indels: 48
DB: 1 Gaps: 5

US-09-647-661-1 (1-900) x Y120_METTH (1-191)
QY 181 CAACTATTAAACGCGCATCTTCAAGATGTTTGACAGCCATTATGAGTTTCTAGT 240
DB 27 GluAlaIleLysThrArgSerIleArgLysTyrGlnAspArg---GluValProGlu 45
QY 241 GAAGATTAGAAATCGCTGAATCGCCAGGCTATCGCAAGCTCTTACACACGCGAG 300
DB 46 GluLeuIleAspLysIleLeuAspAlaAlaMetCysGlyProSerAlaValAspGluArg 65
QY 301 CCATGCGATTGTTGATGTTACTAATAGGATTAAAAACAATTT---GCAGTGCAC 357
DB 66 ProTrpHisPheIleValValArgAsnArgGluMetLeuGluLysIleProGluValHis 85
QY 358 AGCTACTTTTAAATGAAGAAATGATTAAGCGCTTAATCGGTAAATCGGTGATGCTCTTTA 417
DB 86 ProTyr-----GlyAlaMetValLysAspAlaProValAlaIleIleValCysCys--- 102
```





```

GENERAL INFORMATION:
APPLICANT: Anlezark, Gillian M.
APPLICANT: Melton, Roger
APPLICANT: Sherwood, Roger
APPLICANT: Connors, Thomas
APPLICANT: Friedlos, Frank
APPLICANT: Jarman, Michael
APPLICANT: Knox, Richard
APPLICANT: Mauger, Anthony
TITLE OF INVENTION: Bacterial Nitroreductase for the
TITLE OF INVENTION: Reduction of CB 1954 and Analogues Thereof to a Cytotoxic
TITLE OF INVENTION: Form
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spencer, Frank & Schneider
STREET: Suite 300 East, 1100 New York Avenue, N.W.
CITY: Washington, D.C.
COUNTRY: U.S.A.
ZIP: 20005-3955
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/504,047
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/232,018
FILING DATE: 27-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Schneller, John W.
REGISTRATION NUMBER: 26,031
REFERENCE/DOCKET NUMBER: KEMFJ 0017
TELEPHONE: (202) 414-4000
TELEFAX: (202) 414-4040
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Salmonella typhimurium
US-08-504-047-3
Alignment Scores:
Pred. No.: 2,41e-15 Length: 217
Score: 190.50 Matches: 51
Percent Similarity: 48.35% Conservative: 37
Best Local Similarity: 28.02% Mismatches: 55
Query Match: 12.22% Indels: 39
DB: 1 Gaps: 5

US-09-647-661-1 (1-900) x US-08-504-047-3 (1-217)
QY 193 GAGCGCCATCTTCAAGATGTTTCAGACGCAATATGAGTTCTAGTGAAGATTAGAA 252
DB 9 GlnArgTyrSerThrLysAlaPheAspProSerLysLysLeuThrAlaGluAlaAsp 28
QY 253 GAATCGCTGAATCGCAGGCTATCCGAGCTTACACGCGCCATCGCATTTT 312
DB 29 LysileLysThrLeuGlnTyrSerProSerSerThrAsnSerGlnProTrpHisPhe 48
QY 313 GTGATGGTTACTAATAGGATTTTAAAAACAAATTCGATGCGACACG----- 360
DB 49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnTyr 68
```

```

361 TACTTTAATGAGAAATGATTAAAGCGCTTACAGCTTAATGCTGTGATGCTCTTTAAGA 420
DB 69 ThrPheAsnGluArgLysMetLeuAspAlaSerHisValValPheCysAlaLysThr 88
QY 421 CCTAGCGAGTGTGTACACACGCGCCATTACATGCAAAACCTTTACCCGAGTCTTATAAG 480
DB 89 AlaAsnAsp-----AspAlaTrpLeu 95
QY 481 GTTAGAGTATGATC-----CCTTCTTTT 501
DB 96 GluArgValValAspGlnGluAspAlaAspGlyArgPheAlaThrProGluAlaLysAla 115
QY 502 GCTCAAAATGCTTGGCGTGAATTCACACGACGATGCAAGATTA----- 546
DB 116 AlaAsnAspLysGlyArgArgPhePheAlaAspMetHisArgValSerLeuLysAspAsp 135
QY 547 GAAAGCTATATTTTAGAGCAATCTATATCGCTGGGGCAAAATTTGCGATGGCGCGTACG 606
DB 136 HisGlnTrpMetAlaLysValValTyrLeuAsnValGlyAsnPheLeuLeuGlyValAla 155
QY 607 TTAATGGGATGATGATGATGATTTGGAGGCTTTGATCCTTTAAAGGGGTGAAGTT 666
DB 156 AlaMetGlyLeuAspAlaValProIleGluGlyPheAsp-----AlaGluVal 171
QY 667 TTAGAA 672
DB 172 LeuAsp 173

RESULT 4
US-09-087-855-3
; Sequence 3, Application US/09087855
; Patent No. 597065
; GENERAL INFORMATION:
; APPLICANT: Anlezark, Gillian M.
; APPLICANT: Melton, Roger
; APPLICANT: Sherwood, Roger
; APPLICANT: Connors, Thomas
; APPLICANT: Friedlos, Frank
; APPLICANT: Jarman, Michael
; APPLICANT: Knox, Richard
; APPLICANT: Mauger, Anthony
; TITLE OF INVENTION: Bacterial Nitroreductase for the
; TITLE OF INVENTION: Reduction of CB 1954 and Analogues Thereof to a Cytotoxic
; TITLE OF INVENTION: Form
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spencer, Frank & Schneider
; STREET: Suite 300 East, 1100 New York Avenue, N.W.
; CITY: Washington, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3955
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/087,855
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: 08/232,018
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Schneller, John W.
; REGISTRATION NUMBER: 26,031
; REFERENCE/DOCKET NUMBER: KEMFJ 0017
; TELEPHONE: (202) 414-4000
; TELEFAX: (202) 414-4040
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Salmonella typhimurium
; US-08-504-047-3
```





```

Db      29  GlnIleLysThrLeuLeuGlnIntySerProSerSerThrAsnSerGlnProTrpHisPhe 48
QY      313  GTGATGGTTACTAATAAGAGATTTAAAAAACAATAATGCAGTGCACAGC----- 360
Db      49  IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnTyr 68
QY      361  TACTTTTAATGAAGAATGATATAAACGGCTTCAGCGTTAATGCTGGTATGCTCTTTAAGA 420
Db      69  ValPheAsnGluArgLysMetLeuAspAlaSerHisValValPheCysAlaLysThr 88
QY      421  CCTAGCGAGTGTGTACACACAGGCCCATTCATCCAAAAACCTTTAC----- 465
Db      89  Ala-----MetAspAspValTrpLeuLysLeuValVal 99
QY      466  -----CCGAGCTCTTATAAGTTAGATGATC----- 492
Db      100  AspGlnGluAspAlaAspGlyArgPheAlaThrProGluAla----- 113
QY      493  CCTTCITTTTGCTCAATGCTTGGCGTGAGATTCACACACAGCATGCAAGATTAGAA-- 549
Db      114  ---LysAlaAlaAsnAspLysGlyArgLysPheAlaAspMetHisArgLysAspLeu 132
QY      550  -----AGCTATATTTTAGACCAATGCTATATCGCTCTGGGGCAAAATTTGCATG 597
Db      133  HisAspAspAlaGluTrpMetAlaLysGlnValTyLeuAsnValGlyAsnPheLeuLeu 152
QY      598  GCGCTGAGCTTAATGGATGGATAGTAGTTCATTTGGAGCGCTTTGAT 645
Db      153  GlyValAlaAlaLeuGlyLeuAspAlaValProIleGluGlyPheAsp 168

RESULT 7
US-08-640-808-2
; Sequence 2, Application US/08640808
; Patent No. 5958682
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: IMPROVEMENTS RELATING TO CANCER THERAPY
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BFO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,808
; FILING DATE:
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-640-808-2

Alignment Scores:
Pred. No.: 4.32e-15 Length: 217
Score: 186.50 Matches: 49
Percent Similarity: 47.73% Conservative: 35
Best Local Similarity: 27.84% Mismatches: 51
Query Match: 12.09% Indels: 41
DB: 2 Gaps: 5

US-09-647-661-1 (1-900) x US-08-640-808-2 (1-217)
QY      193  GAGCGCATCTTCGCAAGATGTTTGACAGCCATTATGAGTTTCTAGTGAAGATTAGAA 252
Db      9  LysArgHisSerThrLysAlaPheAspAlaSerLysLeuThrProGluGlnAlaGlu 28
QY      253  GAAATGCTGAATCGCCAGCGTATCGCAAGCTCTTACACGCGAGCCATGGCATTTT 312
Db      29  GlnIleLysThrLeuLeuGlnIntySerProSerSerThrAsnSerGlnProTrpHisPhe 48

```



Best Local Similarity: 23.70%				Mismatches:	102
Query Match: 11.80%				Indels:	8
DB: 4				Gaps:	3
 US-09-647-661-1 (1-900) x US-09-540-236-3762 (1-232)					
QY	148	TCATGAAATTTTGGATCAGCAAAAAGAGAACAACCTATTAAAGAGCGCCATTCTTGC	207	:	:::::
Dd	13	ThrMeGlnPheAsnLysAspValLeuAenAlaPheHisTyArgValSerThr	32	:	:::::
QY	208	AAGATCTTGACAGCCATTATCAGTGTTCCTAGTGAAGAAATTAGAGAATAATCGCTGAATC	267	:	:::::
Dd	33	ArgSerTyAspGlyAspLysIleProAlaGluaspPheAsnTyIleLeuGluLeu	52	:	:::::
QY	268	GCAGCATATGCCAAGCTCTTACACAGCAGCCATGGCATTTTGATGGTTACTTAAT	327	:	:::::
Dd	53	GlyArgLeuSerProSerSerValGlySerGluProTrpGlnPheLeuValLeuGlnAsn	72	:	:::::
QY	328	AAGATTAAAAAAAACAAATTCAGTGCACAGCTACTTTAATGAAGAAATGATTAAAGC	387	:	:::::
Dd	73	AlaAspLeuargGlnLysLeuLysProTyrcysTrpGlylleProThrMetGluThrSer	92	:	:::::
QY	388	GCTTCAGCGTTAATGTGTGTATGCTCTTTAAGACCCTAGCGAGTTGTTACACAC--	444	:	:::::
Dd	93	SerHisIleValAlaIleLeuAlaIlysLysAenAlaArgTyArgSerThrProTyPheGly	112	:	:::::
QY	445	CATTACATGC AAAAC-----CTTTACCGGAGTCTTATTAAGGTTAGAGTGATCCCT	495	:	:::::
Dd	113	GluIleMetAspArgGlyLeuThrGlyAspArgAlaLysAlaLeuGluValTy	132	:	:::::
QY	496	TCTTTTGCTCAAATGCTTGCGGTGAGATTCAACACAGCATGCAAGATTAGAAAGCTAT	555	:	:::::
Dd	133	GlnLysPheGlnGluAspIleAlaIleLeuAspSerGluArgSerLeuTyArgTrp	152	:	:::::
QY	556	ATTTAGCAATGCTATATCCCTGTGGGGCAAAATTCATGGCGGTGAGCTTAATGGGA	615	:	:::::
Dd	153	AlaSerLysGlnThrTyrlleAlaLeuAlaAsnMetMetThrGlyAlaAlaMetileGly	172	:	:::::
QY	616	TTGCATAGTTGCATATTTCGAGGCTTTTCATCTTTTAAAGTGGGTGAAGTTTTAGAGAG	675	:	:::::
Dd	173	ValAspSerCysProIleGluGlyPheAlaTyLysGluValasnGlnLeuLeuAlaAsp	192	:	:::::
QY	676	CGT-----ATCAATAAGCCTAAATCGCATCGCTTGATCGCTTTGGGCAAGAG	723	:	:::::
Dd	193	GluGlyLeuPheAspProasnGluTrpAlaValSerValMetValThrPheGlyTyArg	212	:	:::::
QY	724	GTGCGAAGCGAGCGCAAAAATCAAGAAATCA	756	:	:::::
Dd	213	AspLysGluIleLysProLysAlaargLysThr	223	:	:::::
 RESULT 10 US-09-489-039A-13412 ; Sequence 13412, Application US/09489039A ; Patent No. 6610836 ; GENERAL INFORMATION: ; APPLICANT: Gary Breton et. al ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBS ; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS ; CURRENT FILING DATE: 2000-01-27 ; PRIOR APPLICATION NUMBER: US 60/117,747 ; PRIOR FILING DATE: 1999-01-29 ; NUMBER OF SEQ ID NOS: 14342 ; SEQ ID NO 13412 ; LENGTH: 220 ; TYPE: PRS ; ORGANISM: Klebsiella pneumoniae US-09-489-039A-13412					
Alignment Scores:					
Pred.	No.	Length:			
Score:		Matches:			
		49			

Percent Similarity:	47.40%	Conservative:	33
Best Local Similarity:	28.32%	Mismatches:	56
Query Match:	11.58%	Indels:	35
DE:	4	Gaps:	5

US-09-647-661-1 (1-900) x US-09-489-039A-13412 (1-220)

Qy	193	GAGCGCCATCTTGCAGATCTTTGACAGCCATATCAGTTTCTAGTGAAGATTAGAA	252
Db	12	LYSARGTYSERThrLYSAlaPheASPAlaThrLYSLeuThrAlaGlyGluAlaGlu	31
Qy	253	GAAATCGCTGAAATCGCCAGGCTATCGCCAGCTCTTACACACGACGACCATGGCATTTT	312
Db	32	GlnLeuLYSThrLeuLeuGlnTyr-SerProSerSerThrAsnSerGlnProTPrHisPhe	51
Qy	313	GTGATCGTTACTAATAAGCAATTAAAAAACAATAATCCAGTGCACAGC-----	360
Db	52	IleValAlaSerThrASPGLUGLYLeuAlaArgValAlaLYSAlaAlaSerGlyThrTyr	71
Qy	361	TACTTTAATGAAGAAATGATTAAGAGCGCTTCAGCGTTAATGGTGGTATGCTCTTTAAGA	420
Db	72	ValPheAsnGluArgLYSILEuAspAlaSerHisValValValPheCysAla-----	89
Qy	421	CCTPAGCGAGTTGTTACCACACGGCCATTACATGCAAAACCTTTACCGGAGTCTTTAAG	480
Db	90	-----LysThrAlaMetASPAlaTPrLeu	98
Qy	481	GTTAGATGATCCCTTCT-----TTTGTCAAATGCTTGGCGTGAGA	522
Db	99	GlnArgValValASPGLUGLUAlaASPGLYArgPheAlaThrProASPAlaLYSAla	118
Qy	523	TTCAACCAC-----AGCATGCAAAAGATTAGAA-----	549
Db	119	AlaAsnHisLYSGLYArgThrPhePheAlaASPMetHisArgLYSGluLeuLYSASPASP	138
Qy	550	---AGCTATATTTAGCAATGCTATATCCTCTGGGGCAATTTGCATGGCGGTGAGC	606
Db	139	ASPGLnTrpMetAlaLYSGlnValTyrLeuAsnValGlyAsnPheLeuLeuGlyValAla	158
Qy	607	TTAATGGGATTCGATAGTTGCATTATTGGAGCGTTTGAT	645
Db	159	AlaMetGlyLeuASPAlaValProIleGluGlyValASP	171

RESULT 11  
US-09-328-352-6977  
; Sequence 6977, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETO  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6977  
; LENGTH: 218  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6977

Alignment Scores:	1,67e-13	Length:	218
Pred. No.:	176.00	Matches:	49
Score:	45.92%	Conservative:	41
Percent Similarity:	25.00%	Mismatches:	86
Best Local Similarity:	11.29%	Indels:	20
Query Match:	4	Gaps:	5
DE:			

US-09-647-661-1 (1-900) x US-09-328-352-6977 (1-218)

Qy	184	CTATTAAAC-----GAGCGCCATCTTGCAGATGTTTGAACGCCATTATGAGTTT	234

```

Db      4 LeuLeuAsnThrVallySerArgTyrThrThrLysAlaTyrAspProGluLysIle 23
Qy      235 TCAGTGAAGAAATGAAGAAATCGCTGAATCGCCAGGCTATCGCAAGCTCTACAAAC 294
Db      24 ProGlnGluLysPheAsnLysLeuLeuGluLeuLeuArgPheThrProSerValAsn 43
Qy      295 ACCGACGATGGCATTTTGATGGTGTACTAATAAGGATTTAAATAACAAATTCAGTG 354
Db      44 IleGlnProTrpHisPheLeuValAlaAspAsnProThrAlaLysGluArgIleAlaLys 63
Qy      355 -----CACAGCTACTTAAATGAAGAATGATTAAGCGCTTCAGCGTTAATG 402
Db      64 AlaLeuThrGlyArgTyrAlaTyrAsnAlaProLysValLeuLysSerHisThrLeu 83
Qy      403 GTGTAATGCTCTTAAGACCTAGGAGTGTGTACACACAGCGCCATACATGCAAAACCTT 462
Db      84 ValPheCysThr-----ArgThrAspIleSerPro-----GluTyrLeuAsnGlnLeu 99
Qy      463 TACCGGAGCTTATAAGGTTAGAGTATCCCTCTTCTGCTCAATGCTTGGCGTGAGA 522
Db      100 LeuGluGlnAspLeuSerGlyArgPheLysAspGluLysAlaLysLeuGlyGlnLys 119
Qy      523 -----TTCAACACACAGCATGCCAAAGATTAGAAAGCTAT 555
Db      120 AspThrArgHisGlyTyrValGluPheTyrArgAsnGluGlnLysAsnLeuPheGlyTrp 139
Qy      556 ATTTTAGAGCAATGCTATATCGCTGTGGGCAAAATTTGCAATGGCGCTGAGCTTAATCGGA 615
Db      140 MetGluAsnGlnThrPheIleAlaLeuGlyGlnLeuLeuPheAlaGlyLeuGluGly 159
Qy      616 TTGATAGTTCATTATGAGGCTTTGATCCCTTTAAAGTGGGTGAGTTTAGAAGAG 675
Db      160 IleAspAlaThrProMetGlyGlyPheAspGluAsnValLeuAsnGluGluPheGlyLeu 179
Qy      676 CGTATCAATAAGCCATAAATCGCATGCTGTATCGCTTTGGGCAAGAGG 723
Db      180 LysGluLysGlyLeuArgSerSerValIleValSerLeuGlyTyrArg 195

RESULT 12
US-09-107-532A-5919
; Sequence 5919, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Dereke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007

```

```

; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5919:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...218
; SEQUENCE DESCRIPTION: SEQ ID NO: 5919:
US-09-107-532A-5919

Alignment Scores:
Pred. No.: 1,28-10 Length: 218
Score: 153.50 Matches: 49
Percent Similarity: 45.00% Conservative: 50
Best Local Similarity: 22.27% Mismatches: 84
Query Match: 9.85% Indels: 37
DB: 4 Gaps: 6

US-09-647-661-1 (1-900) x US-09-107-532A-5919 (1-218)

Qy      166 CAGGAAAAAAGAGACAACTA-----TTA 189
Db      1 GlnGlnLysArgLysArgLeuIleMetGluGlnLysAsnAspPheSerThrIleMet 20
Qy      190 AACGAGCGCCATTTGCAAGATGTTGACAGCCATATGAGTTTCTAGTGAAGAATTA 249
Db      21 ThrGlyArgArgSerValArgValThrIleProHisGluGluMet 40
Qy      250 GAAGAAATCGCTGAATCGCCAGGCTATCGCAAGCTTTACAAACAGCCAGCTATGAT 309
Db      41 LeuGluMetIleLysGluAlaThrThrAlaProSerSerValAsnMetGlnProTrpArg 60
Qy      310 TTTGTGATGGTTACTAATAAGGATTTAAAAACAATTTGAGTGCAGTCAGCTACTTAAT 369
Db      61 PheValValAlaGluSerGluGluAlaLysGluLeuLeuArgProLeuIleArgPheAsn 80
Qy      370 GAAGAAATGATTTAAAGCGCTTCAGCGTTAATGGTGTA-----TGCTCT 414
Db      81 ThrArgGlnAsnAspThrSerSerAlaMetValMetIlePheGlyAspMetGlnCysTyr 100
Qy      415 TTAAGACCTAGGAGCTGTTTACCACACCGCCATACATGCAAAACCTTTACCGGAGTCT 474
Db      101 -----GluTyrGlyGluGluIleTyrAsnGlnAla 110
Qy      475 TATAAGGTTAGAGTATCCCT-----TCTTTTGCTCAAAATGCTTGGC 516
Db      111 TyrGluSerGlyLysMetProLysGluValArgAspGlnGlnLeuAlaAlaIlePro 130
Qy      517 GTGAGATTCACACACAGCATGCAAGATTAAGACCTATATTTAGACCAATGTATATC 576
Db      131 TyrTyrLysSerPheSerArgGluMetAsnAspValValLysValAspSerSerLeu 150
Qy      577 GCTGTGGGCAATTTGCGATGGCGTGAAGTTAATGGATGGATGATGATGATTTATGGA 636
Db      151 AlaAlaMetGlnPheMetLeuValAlaArgAspHisGlyTyrGluThrAsnProIleGly 170
Qy      637 GCGTTTGTATCCTTTAAAAAGTGGTGAAAGTT-----TTAGAGAGCGGTATCAATAGCCT 690
Db      171 GlyPheGluAlaAspGlnLeuAlaGluAlaPheGlyLeuAspLysAspArgTyrValPro 190
Qy      691 AAAATCGCATGCTTGTATCGCTTTGGCAAGAGGTTGGCAGAGCGCAGCAAAATCAAGA 750
Db      191 ValIle-----IleLeuSerIleGlyLysAlaValGluGluTyrGluSerValArg 208

RESULT 13
US-08-412-108-4
; Sequence 4, Application US/08412108

```



Wed Apr 21 10:19:02 2004

[illegible]

RESULT 15  
US-09-352-991A-23965  
; Sequence 22965, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```

, TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
, FILE REFERENCE: 107196.136
, CURRENT APPLICATION NUMBER: US/09/252,991A
, CURRENT FILING DATE: 1999-02-18
, PRIOR APPLICATION NUMBER: US 60/074,788
, PRIOR FILING DATE: 1998-02-18
, PRIOR APPLICATION NUMBER: US 60/094,190
, PRIOR FILING DATE: 1998-07-27
, NUMBER OF SEQ ID NOS: 33142
, SEQ ID NO 22965
, LENGTH: 221
, TYPE: PRT
, ORGANISM: Pseudomonas aeruginosa
, US-09-252-991A-22965

```

Alignment Scores:	
Pred. No.:	7,23e-09
Score:	139.50
Percent Similarity:	46.50%
Best Local Similarity:	25.00%
Query Match:	8.95%
DB:	4
Length:	221
Matches:	50
Conservative:	43
Mismatches:	78
Indels:	29
Gaps:	8

US-09-647-661-1 (1-900) x US-09-252-991A-22965 (1-221)

Qy	187	TTAAACAGCAGCGCCATCTTGCACAGATGTTTTGACAGCATATTAGAGTTTCTTAGTGAGAA	246
Db	30	VAlArGSerAArgAlAlleIleGlytYrAspSerSerpHeSerLeuthrArgGluGlu	49
Qy	247	TTAGAAGAAAATCGCTGAATCCGCCAGGCTATCCGCCAAGCTCTTACAACACGACGCCCATGG	306
Db	50	LysAspHisLeuLeuAspLeuAlaLeuHiealAprSerAlaPheAsnLeuGlnHisVal	69
Qy	307	CATTTTGTGATGGTTACTTAATAAGGATTTAAAAAACAAATTGCATGTCACAGCTACTTT	366
Db	70	ArgLeuValGluValSerAspProGlnLeuArgValGlnLeuArgGluValAlatrp---	88
Qy	367	AATGAAGAAAATGAATAAAGCGCTTCACGCTTAATGGTGTATGCTCTCTTAAAGACCTAGC	426
Db	89	AspGlnAlaGlnValThrAspAlaAlaMetLeuValValCys-----Ala	104
Qy	427	GAGTTGTTCACACAGCGCCATTACATCAAAACCTTTAC-----CGGAGTCTTAT	477
Db	105	GlnLeuAppSerTrpGluArgAsnAlaGlnArgValTrpAspGlyAlaProGluAlaVal	124
Qy	478	AAGGTTAGATGATCCCTCTTTTGCTCAAATGCTTGGCGTG-----AGA	522
Db	125	Gln-----AlapHeMetAlaGlyAlaIleAspThrtYrTyArG	137
Qy	523	TTCNACCACGATCGCAAGATTGAAGCTATATTATGAGCAATGCTATATCGCTGG	582
Db	138	GlyLysProGlnValGlnArgaspGluAla-----MetArgSerCysglyLeuLeuAla	155
Qy	583	GGGCAAAATTGCATGGCGGTGAGCTTAATGGATGGATAGTTGCATATTATGGAGGCTTT	642
Db	156	GlnThrLeuMetLeuValAlaArgGlyGlnGlyLeuAspSerCysProMetAspGlyPhe	175
Qy	643	GATCCTTTAAAAGTCGGTGAAGTTTATAGAGCGGTATCAATAAGCTAAA-----	693
Db	176	App-----PheaspAlavalAlaargLeuIleasnLeuProaspaenHisval	191
Qy	694	ATCGCATGCTTGATCGCTTTGGCAAGAGGCTGGCAGACGAGCCAAATAATCAAGAAA	753
Db	192	IleGlnLeuMetValAlaValGlyLysLysAlaValGluProTrpProArgSerGlyLys	211

Search completed: April 20, 2004, 19:12:54  
Job time : 33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 20, 2004, 19:12:00 ; Search time 59 Seconds

(without alignments)  
8620.093 Million cell updates/sec

Title: US-09-647-661-1

Perfect score: 1559  
Sequence: 1 tgcagaattttacagagagc.....gggggtgttttcaagcggtttc 900

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DRV=xlp  
-Q/cgr2\_1/USFTO.spool\_p/US09647661/runat\_20042004\_135347\_2622/app\_query.fasta\_1.1095  
-DB=A\_Geneseq\_29Jan04 -QFWT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOKCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09647661 @CGN 1.1 81 @runat\_20042004\_135347\_2622 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -YGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1072	68.8	210	2	Aaw98657 H. pylori
2	1071.5	68.7	209	2	Aay43259 H. pylori
3	1067	68.4	210	3	Aay90347 H. pylori
4	258	16.5	222	4	Aag82819 S. epider
5	258	16.5	222	4	Aag82690 S. epider
6	258	16.5	226	5	Abp39622 Staphyloc
7	246	15.8	284	2	Aaw98436 H. pylori
8	234	15.0	240	6	Abm72562 Staphyloc
9	210.5	13.5	221	6	Abp78332 N. gonorr
10	204.5	13.1	190	3	Aay90339 H. influe

11	198.5	12.7	201	3	AAY90344	Aay90344 C. jejuni
12	192.5	12.3	217	2	AAW13084	Aaw13084 Escherich
13	192.5	12.3	217	2	AAW13083	Aaw13083 Escherich
14	192.5	12.3	217	6	ABP70750	Abp70750 Mutant E.
15	190.5	12.2	217	6	ABP70759	Abp70759 Mutant E.
16	189.5	12.2	217	6	ABP70754	Abp70754 Mutant E.
17	189.5	12.2	217	6	ABP70755	Abp70755 Mutant E.
18	188.5	12.1	217	2	AAR34723	Aar34723 E. coli B
19	188.5	12.1	217	2	AAR3968	Aar3968 Nitroredu
20	188.5	12.1	217	2	AAR76580	Aar76580 E. coli n
21	188.5	12.1	217	4	AAU29347	Aau29347 Novel mar
22	188.5	12.1	217	6	ABP70738	Abp70738 Wild-type
23	188.5	12.1	217	6	ABP70751	Abp70751 Mutant E.
24	188.5	12.1	614	4	ABG34803	Abg24803 Novel hum
25	186.5	12.0	217	6	ABP70752	Abp70752 Mutant E.
26	185.5	11.9	202	3	AAI90336	Aay90336 B. subtil
27	185.5	11.9	217	2	AAR95880	Aar95880 Bacterium
28	184.5	11.8	217	6	ABP70753	Abp70753 Mutant E.
29	182	11.7	221	4	AAU03119	Aau03119 Streptoco
30	182	11.7	221	5	ABP27441	Abp27441 Streptoco
31	181.5	11.6	217	6	ABP70756	Abp70756 Mutant E.
32	181.5	11.6	217	6	ABP70757	Abp70757 Mutant E.
33	181.5	11.6	217	6	ABP70758	Abp70758 Mutant E.
34	176	11.3	218	6	ADA35690	Ada35690 Acinetoba
35	175.5	11.3	248	5	ABP27440	Abp27440 Streptoco
36	172.5	11.1	209	3	AAI90335	Aay90335 B. subtil
37	172	11.0	224	5	ABP30417	Abp30417 Streptoco
38	165	10.6	180	3	AAI90341	Aay90341 Synechocy
39	160	10.3	201	6	ABU0998	Abu0998 S. pneumo
40	159	10.2	210	5	ABH43245	Abh43245 Listeria
41	153.5	9.8	218	7	ADC96292	Adc96292 E. faeciu
42	153.5	9.8	356	5	ABB54815	Abb54815 Lactococc
43	153	9.7	174	3	AAI90343	Aay90343 A. fulgid
44	151.5	9.7	202	5	ABU50992	Abu50992 Helicobac
45	150	9.6	198	3	AAI81711	Aay81711 Streptoco

#### ALIGNMENTS

RESULT 1

AAW98657

ID AAW98657 standard; protein; 210 AA.

XX AC AAW98657;

XX DT 31-MAR-1999 (first entry)

XX DE H. pylori GHPO 455 protein.

XX KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
peptic ulcer disease.

XX OS Helicobacter pylori.

XX PN WO9843478-A1.

XX PD 08-OCT-1998.

XX PF 01-APR-1998; 98WO-US006371.

XX PR 01-APR-1997; 97US-00833457.

XX PR 24-JUN-1997; 97US-00881227.

XX PR 29-JUL-1997; 97US-00902615.

XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX14376.

XX New isolated Helicobacter polynucleotides - used to develop products for

PT the diagnosis, prevention and treatment of Helicobacter infections and  
PT gastrointestinal diseases.  
XX  
PS Claim 8; Page 1347-1348; 2054pp; English.  
XX

CC This sequence represents a Helicobacter pylori GHPD protein of the  
CC invention. The polypeptides can be used for preventing or treating  
CC Helicobacter infections, and gastroduodenal diseases associated with  
CC these infections, including acute, chronic, and atrophic gastritis, and  
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
CC used for the production of antibodies. The products can also be used for  
CC detection and diagnosis  
XX  
SQ Sequence 210 AA;

Alignment Scores:  
Pred. No.: 5.4e-117 Length: 210  
Score: 1072.00 Matches: 207  
Percent Similarity: 99.05% Conservative: 1  
Best Local Similarity: 98.57% Mismatches: 2  
Query Match: 68.76% Indels: 0  
DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x AAW98657 (1-210)

QY	151	ATGAAATTTTGGATCGAGAAAGAGACAACTATTAAACGAGCGCATTTCTTGCAG	210
DB	1	MetLysPheLeuAspGlnGluLysArgGlnLeuLeuAsnGluArgHisSerCysLys	20
QY	211	ATGTTTACAGCCATTATGATTTCTAGTGAAGAATTAGAAGAAATCGCTGAAATCGCG	270
DB	21	MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluGluLeuLeuLeuA	40
QY	271	AGGTATCGCAGCTCTTACACAGCAGCCATGCGATTTTGTGATGTTACTAATAG	330
DB	41	ArgLeuSerProSerSerTyrAsnThrGlnProTrpHisPheValMetValThrAspLys	60
QY	331	GATTTAAAAACAAATTCGAGTGCACAGCTACTTTAAATCAAGAAATGATTAAGCGCT	390
DB	61	AspLeuLysLysGlnIleAlaAlaHisSerTyrPheAsnGluGluLeuLeuLeuA	80
QY	391	TCAGCTTAATGGTGTATGCTCTTTAAGACCTAGCAGTGTGTACACAGCGCATTC	450
DB	81	SerAlaLeuMetValValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr	100
QY	451	ATGCAAACTTTTACCCGGAGTCTTATAAGTTAGATGATCCCTTTCTTTGCTCAAATG	510
DB	101	MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet	120
QY	511	CTTGCGCTGAGATTCAACACAGCATGCCAAAGATTAGAAGCTATATTTAGAGCAATGC	570
DB	121	LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGlnCys	140
QY	571	TATATCCCTGTGGGCAAAATTTGATGGCGTGCAGCTTAATGGATTGGATGTCATT	630
DB	141	TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle	160
QY	631	ATTGGAGCTTTGATCCCTTTAAAGTGGGTGAAGTTTGAAGAGCGGTATCAATAAGCCT	690
DB	161	IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro	180
QY	691	AAATCCAGCTGTGATCGCTTTGGGCAAGAGGTGGCAGAGCGCAAGCAAAATCAAGA	750
DB	181	LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg	200
QY	751	AAATCAAAAGTGTGATGCGATTACTTGTGTTG 780	
DB	201	LysSerLysValAspAlaIleThrTrpLeu 210	

RESULT 2  
AAV43259  
ID AAV43259 standard; protein; 209 AA.  
XX

AC AAV43259;  
XX  
DT 17-JAN-2000 (first entry)  
XX  
DE H. pylori nitroreductase, rdxA, protein sequence.  
XX

XX Nitroreductase; rdxA; NADPH; nicotinamide-adenine dinucleotide phosphate;  
KW electron donor; prodrug conversion; cytotoxic compound generation;  
KW growth inhibitor; tumour cell; neoplastic cell.  
XX  
OS Helicobacter pylori.  
XX  
PN WO9951270-A1.  
XX  
PD 14-OCT-1999.  
XX  
PF 06-APR-1999; 99WO-US007546.  
XX  
PR 06-APR-1998; 98US-0080917P.  
PR 14-APR-1998; 98US-0081778P.  
XX  
PA (UYDA-) UNIV DALHOUSIE.  
XX

PI Goodwin A, Hoffman PS;  
XX  
DR WPI; 1999-620172/53.  
DR N-PSDB; AAZ31668.  
XX  
PT New nitroreductase that can convert aromatic nitro prodrugs to cytotoxic  
PT compounds, e.g. for killing or inhibiting cancer cells and microbes.  
XX  
PS Disclosure; Fig 1; 40pp; English.  
XX

CC This sequence is the helicobacter pylori nitroreductase, designated rdxA,  
CC of the invention. The nitroreductase has an isoelectric point (pI) over 6  
CC ; contains at least two Cys; has a preference for NADPH (nicotinamide-  
CC adenine dinucleotide phosphate) as electron donor; and can convert a  
CC prodrug to one or more cytotoxic compounds. The nitroreductase is used to  
CC generate cytotoxic compounds for killing or inhibiting growth of targeted  
CC cells (bacteria, (retro) viruses, fungi, immune system cells, tumour  
CC cells, neoplastic or other diseased cells, or cells from a particular  
CC tissue or organ), especially where conjugated with a targeting compound.  
CC The nitroreductase can also be used to select against cells that express  
CC a functional nitroreductase-encoding gene, e.g. to detect Helicobacter  
CC pylori that are resistant to metronidazole and to identify substrates for  
CC nitroreductase  
XX  
SQ Sequence 209 AA;

Alignment Scores:  
Pred. No.: 6.18e-117 Length: 209  
Score: 1071.50 Matches: 209  
Percent Similarity: 99.52% Conservative: 0  
Best Local Similarity: 99.52% Mismatches: 0  
Query Match: 68.73% Indels: 1  
DB: 2 Gaps: 1

US-09-647-661-1 (1-900) x AAV43259 (1-209)

QY	151	ATGAAATTTTGGATCGAGAAAGAGACAACTATTAAACGAGCGCATTTCTTGCAG	210
DB	1	MetLysPheLeuAspGlnGluLysArgGlnLeuLeuAsnGluArgHisSerCysLys	20
QY	211	ATGTTTACAGCCATTATGATTTCTAGTGAAGAATTAGAAGAAATCGCTGAAATCGCC	270
DB	21	MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluGluLeuLeuLeuA	40
QY	271	AGGTATCGCAGCTCTTACACAGCAGCCATGCGATTTTGTGATGTTACTAATAG	330
DB	41	ArgLeuSerProSerSerTyrAsnThr---ProTrpHisPheValMetValThrAsnLys	59
QY	331	GATTTAAAAACAAATTCGAGTGCACAGCTACTTTAATCAAGAAATGATTAAGCGCT	390





OS Staphylococcus epidermidis.  
XX WO200134809-A2.  
XX 17-MAY-2001.  
XX 09-NOV-2000; 2000WO-US030782.  
XX 09-NOV-1999; 99US-0164258P.  
XX (GLAX ) GLAXO GROUP LTD.  
XX Kimmerly WJ;  
XX WPI; 2001-316495/33.  
XX N-PSDB; AAH53669.  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
XX useful for vaccinating against infections, e.g. endocarditis.  
XX Claim 18; Page 712; 2188pp; English.  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
XX and (II) can have antibacterial activity and therefore can be used in  
XX vaccination. The nucleic acids (I) may be used to produce the S.  
XX epidermidis polypeptides (II) via the production of vectors containing  
XX them which are used to produce hosts cells which express the  
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
XX used to vaccinate subjects and to raise antibodies against the bacteria.  
XX The polypeptides may also be used to assay for other inhibitors of their  
XX activity and therefore identify compounds that may be used for the  
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA  
XX polynucleotide sequences from the present invention. AAH55091 to AAH55098  
XX represent oligonucleotide sequences and primers which are used in the  
XX exemplification of the present invention. N.B. The present invention  
XX specifically claims all the polynucleotide sequences given in the  
XX sequence listing of the present specification, however the sequence  
XX listing only goes up to SEQ ID NO:4454 so even though sequences are given  
XX in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
XX for SEQ ID NO:4455 to 4464  
XX Sequence 222 AA;  
SQ

Alignment Scores:  
Pred. No.: 4,7e-21 Length: 222  
Score: 258.00 Matches: 63  
Percent Similarity: 53.11% Conservative: 48  
Best Local Similarity: 30.14% Mismatches: 80  
Query Match: 16.55% Indels: 18  
DB: 4 Gaps: 6

US-09-647-661-1 (1-900) x AAG82819 (1-222)

QY 196 CGCATTCTTGCAGATGTTTGCAGCCATTATGATTTTCTAGTGAAGAATTAGAGAA 255  
DB 17 ArgHisAlaThrLysGluPheAspProThrLysLysIleSerAspGluAspPheAsnThr 36  
QY 256 ATGCTGAATGCCAGGTATGCCAGCTTACACACCCAGCCATGGATTTGTG 315  
DB 37 IleLeuGluThrGlyArgLeuSerProSerSerLeuGlyLeuGluProTrpHisPheVal 56  
QY 316 ATGTTTACTAATAAGATTATAAATAACAAATTCAGTCAGTCAGCTACTTTAATGAAGAA 375  
DB 57 ValValGlnAsnLysGluLeuArgGluLysLeuLysAlaLysSerTrpGlyAlaGlnLys 76  
QY 376 ATGATTAAAGCCTTCAGGTTATGGTGGTATGCTCTTTAAGCATTAGCAGTTGTTA 435  
DB 77 GlnLeuAspThrAlaSerHisPheValLeuIlePheAlaArgLys-----AsnValThr 94  
QY 436 CCACAGCGGCATTACATGCAAAACCTTTAC-----CCGAGCTCTTATAAGCTTAGAGTG 489

Db 95 AlaHisThrAspTyrValGlnHisLeuLeuArgGlyValLysLysTyrGluGluSerThr 114  
Qy 490 ATCCCTTCT-----TTTGCTCAAAATGCTTGGCGTGAGATTCAAC 528  
Db 115 IleProAlaValGluAsnLysPheAspPheGlnGluSerPheHisIleAlaAspAsn 134  
Qy 529 CACAGCATGCAAGATTAGAACTATATTTTAGCAATGCTATATCGCTGTGGGGCAA 588  
Db 135 Glu-----ArgThrLeuTyrAspTrpAlaSerLysGlnThrTyrIleAlaLeuAlaAsn 152  
Qy 589 ATTGCAATGGCGTGAGCTTAATGGAATGATGATGATGATGATGATGATGATGATGAT 648  
Db 153 MetMetThrSerAlaAlaLeuLeuGlyIleAspSerCysProIleGluGlyPheAspLeu 172  
Qy 649 TTAAGAGTGGTGAAGCTTTTAGAAGAGCT-----ATCAATAGCCATAAATC 696  
Db 173 AspLysValThrGluIleLeuSerAspGluGlyValLeuAspThrGluGlnPheGlyIle 192  
Qy 697 GCATGCTGTATCGCTTTGGGCAAGAGGCTGGCAGAACGAGCCAA---AAATCAAGAAAA 753  
Db 193 SerValMetValGlyPheGlyTyrArgAlaGlnGluProLysHisGlyLysValArgGln 212  
Qy 754 TCAAAAGTTGATCGATTACTTCTGTTG 780  
Db 213 AsnGluAspAspIleIleSerTrpIle 221

RESULT 5  
AAG82690  
ID AAG82690 standard; protein; 222 AA.  
AC AAG82690;  
XX  
XX 03-SEP-2001 (first entry)  
DT  
XX  
XX S. epidermidis open reading frame protein sequence SEQ ID NO:2474.  
XX Staphylococcus epidermidis S81 strain; infection; diagnosis; vaccination;  
XX endocarditis.  
XX  
XX Staphylococcus epidermidis.  
XX  
XX WO200134809-A2.  
XX  
XX 17-MAY-2001.  
XX  
XX 09-NOV-2000; 2000WO-US030782.  
XX  
XX 09-NOV-1999; 99US-0164258P.  
XX (GLAX ) GLAXO GROUP LTD.  
XX  
XX Kimmerly WJ;  
XX  
XX WPI; 2001-316495/33.  
XX N-PSDB; AAH53540.  
XX  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
XX useful for vaccinating against infections, e.g. endocarditis.  
XX  
XX Claim 18; Page 652; 2188pp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
XX and (II) can have antibacterial activity and therefore can be used in  
XX vaccination. The nucleic acids (I) may be used to produce the S.  
XX epidermidis polypeptides (II) via the production of vectors containing  
XX them which are used to produce hosts cells which express the  
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
XX used to vaccinate subjects and to raise antibodies against the bacteria.  
XX The polypeptides may also be used to assay for other inhibitors of their  
XX activity and therefore identify compounds that may be used for the  
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA

CC polynucleotide sequences from the present invention. AAH55091 to AAH55098  
 CC represent oligonucleotide sequences and primers which are used in the  
 CC exemplification of the present invention. N.B. The present invention  
 CC specifically claims all the polynucleotide sequences given in the  
 CC sequence listing of the present specification, however the sequence  
 CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
 CC in the disclosure for SEQ ID NO:4455 to 4472, no sequences are present  
 CC for SEQ ID NO:4455 to 4464  
 XX  
 XX  
 XX Sequence 222 AA;

# Alignment Scores:

Pred. No.: 4,73e-21 Length: 222  
 Score: 258.00 Matches: 63  
 Percent Similarity: 53.11% Conservative: 48  
 Best Local Similarity: 30.14% Mismatches: 80  
 Query Match: 16.55% Indels: 18  
 DB: Gaps: 6

US-09-647-661-1 (1-900) x AAG82690 (1-222)

QY 196 CGCCATCTTTCAGAGATGTTTACAGCCATTATGAGTTTCTAGTGAAGAATTAGAAGAA 255  
 Db 17 ArgHisAlaThrLysGluPheAspProThrLysLysIleSerAspGluAspPheAsnThr 36  
 QY 256 ATCGCTGAATGCCAGGCTATCGCAAGCTCTTACACACGACGACCATGGCATTTGTG 315  
 Db 37 IleLeuGluThrGlyArgLeuSerProSerSerLeuGlyLeuGluProThrHisPheVal 56  
 QY 316 ATGTTTACTAATAAGATTTAAACAAATTCGACGACGACGACGACGACGACGACGACG 375  
 Db 57 ValValGlnAsnLysGluLeuArgGlyLysLeuLysAlaTyrSerTrpGlyAlaGlnLys 76  
 QY 376 ATGATTAAAGCGCTTCAGCGTTTAATGGTGTATGCTTTAAGACCTAGCGAGTTGTTA 435  
 Db 77 GlnLeuAspThrAlaSerHisPheValLeuIlePheAlaArgLys-----AsnValThr 94  
 QY 436 CCACAGCGCCATTACATGCAAAACCTTTAC-----CCGAGCTCTTATAGGTTAGAGTG 489  
 Db 95 AlaHisThrAspTyrValGlnHisLeuLeuArgGlyValLysLysTyrGluGluSerThr 114  
 QY 490 ATCCCTTCT-----TTTGCTCAAATGCTTGGCGTGAGATTCAAC 528  
 Db 115 IleProAlaValGluAsnLysPheAspPheGlnGluSerPheHisIleAlaAspAsn 134  
 QY 529 CACAGCATGCAAGATTAGAACTATATTTAGACCAATGCTATATCCCTGTGGGCAA 588  
 Db 135 Glu-----ArgThrLeuTyrAspTrpAlaSerLysGlnThrTyrIleAlaLeuAlaAsn 152  
 QY 589 ATTTCATGCGCTGAGCTTAATGGGATTGGATTGCTATTATGCGGCTTGATCCT 648  
 Db 153 MetMetThrSerAlaAlaLeuLeuGlyIleAspSerCysProIleGluGlyPheAspLeu 172  
 QY 649 TTAAGAGTGGTGAAGTTTAGAAGACGCT-----ATCAATAGCCTAAATC 696  
 Db 173 AspLysValThrGluIleLeuSerAspGluGlyValLeuAspThrGluGlnPheGlyIle 192  
 QY 697 GCATGCTTGATCGCTTGGGCAAGAGGTGGCAGAGCGAGCCAA--AAATCAAGAA 753  
 Db 193 SerValMetValGlyPheGlyTyrArgAlaGlnGluProLysHisGlyLysValArgGln 212  
 QY 754 TCAAAAGTTCATGCGATTACTTGTGTTG 780  
 Db 213 AsnGluAspAspIleIleSerTrpIle 221

RESULT 6  
 ABP39622

ID ABP39622 standard; protein; 226 AA.

XX

AC ABP39622;

XX

DT 24-JUL-2002 (first entry)

XX

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4467.  
 XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 XX Staphylococcus epidermidis.  
 OS  
 XX  
 PN US6380370-B1.  
 XX  
 XX 30-APR-2002.  
 XX  
 XX 13-AUG-1998; 98US-00134001.  
 PF  
 XX 14-AUG-1997; 97US-0055779P.  
 PR 08-NOV-1997; 97US-0064964P.  
 XX  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX  
 XX Doucette-Stamm LA, Bush D;  
 PI  
 XX WPI; 2002-381255/41.  
 DR N-PSDB; ABN92167.  
 DR  
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermis  
 PT polypeptide, useful for diagnosing and treating bacterial infections.  
 XX  
 XX Disclosure; SEQ ID NO 4467; 267pp; English.  
 PS  
 XX  
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life cycle  
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX  
 XX Sequence 226 AA;

Alignment Scores:  
 Pred. No.: 4,73e-21 Length: 226  
 Score: 258.00 Matches: 63  
 Percent Similarity: 53.11% Conservative: 48  
 Best Local Similarity: 30.14% Mismatches: 80  
 Query Match: 16.55% Indels: 18  
 DB: Gaps: 6

US-09-647-661-1 (1-900) x ABP39622 (1-226)

QY 196 CGCCATCTTTCAGAGATGTTTACAGCCATTATGAGTTTCTAGTGAAGAATTAGAAGAA 255  
 Db 21 ArgHisAlaThrLysGluPheAspProThrLysLysIleSerAspGluAspPheAsnThr 40  
 QY 256 ATCGCTGAATGCCAGGCTATCGCAAGCTCTTACACACGACGACCATGGCATTTGTG 315  
 Db 41 IleLeuGluThrGlyArgLeuSerProSerSerLeuGlyLeuGluProThrHisPheVal 60  
 QY 316 ATGTTTACTAATAAGATTTAAACAAATTCGACGACGACGACGACGACGACGACGACG 375  
 Db 61 ValValGlnAsnLysGluLeuArgGlyLysLeuLysAlaTyrSerTrpGlyAlaGlnLys 80  
 QY 376 ATGATTAAAGCGCTTTCAGCGTTTAATGGTGTATGCTTTAAGACCTAGCGAGTTGTTA 435  
 Db 81 GlnLeuAspThrAlaSerHisPheValLeuIlePheAlaArgLys-----AsnValThr 98  
 QY 436 CCACAGCGCCATTACATGCAAAACCTTTAC-----CCGAGCTCTTATAGGTTAGAGTG 489  
 Db 99 AlaHisThrAspTyrValGlnHisLeuLeuArgGlyValLysLysTyrGluGluSerThr 118  
 QY 490 ATCCCTTCT-----TTTGCTCAAATGCTTGGCGTGAGATTCAAC 528

Db 119 IleProAlaValGluAsnLysPheAspPheGlnGluSerPheHisIleAlaAspAsn 138  
QY 529 CACAGCATGCAAGATTAGAAAGCTATATTTAGAGCAATGCTATATCGTGTGGGCGAA 588  
Db 139 Glu-----ArgThrLeuTyrAspTrpAlaSerLysGlnThrTyrIleAlaLeuAlaAsn 156  
QY 589 ATTTCATGCGCGTAGACTTAATGCGATTGAGTAGTTCATTTAGAGGCTTTGATCCT 648  
Db 157 MetMetThrSerAlaAlaLeuLeuGlyIleAspSerCysProIleGluGlyPheAspLeu 176  
QY 649 TTAAGCTGGGTGAAGTTTGAAGACCGT-----ATCATAAGCTTAATC 696  
Db 177 AspLysValThrGluIleLeuSerAspGluGlyValLeuAspThrGluGlnPheGlyIle 196  
QY 697 GCATGCTTGATCGCTTTGGGCAAGAGGTGGCGAAGCGAGCCAA---AAATCAAGAAAA 753  
Db 197 SerValMetValGlyPheGlyTyrArgAlaGlnGluProLysHisGlyLysValArgGln 216  
QY 754 TCAGAACTTGATCGGATTACTTGTTG 780  
Db 217 AsnGluAspAspIleIleSerTrpIle 225

RESULT 7  
AAW98436  
ID AAW98436 standard; protein; 284 AA.  
XX AC AAW98436;  
XX AC AAW98436;  
XX 31-MAR-1999 (first entry)  
XX H. pylori GHPO 456 protein.  
XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
XX peptic ulcer disease.  
XX Helicobacter pylori.  
XX WO9843478-A1.  
XX 08-OCT-1998.  
XX 01-APR-1998; 98WO-US006371.  
XX 01-APR-1997; 97US-00833457.  
XX 24-JUN-1997; 97US-00881227.  
XX 29-JUL-1997; 97US-00902615.  
XX (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;  
XX WPI; 1998-542293/46.  
XX N-PSDB; AAX14155.  
XX New isolated Helicobacter polynucleotides - used to develop products for  
XX the diagnosis, prevention and treatment of Helicobacter infections and  
XX gastrointestinal diseases.  
XX Claim 8; Page 743-744; 2054pp; English.  
XX This sequence represents a Helicobacter pylori GHPO protein of the  
XX invention. The polypeptides can be used for preventing or treating  
XX Helicobacter infections, and gastroduodenal diseases associated with  
XX these infections, including acute, chronic, and atrophic gastritis, and  
XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
XX used for the production of antibodies. The products can also be used for  
XX detection and diagnosis

SQ Sequence 284 AA;  
Alignment Scores:  
Pred. No.: 1.32e-19 Length: 284

Score: 246.00 Matches: 49  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.00% Mismatches: 0  
Query Match: 15.78% Indels: 0  
DB: 2 Gaps: 0

US-09-647-661-1 (1-900) x AAW98436 (1-284)

QY 2 GCAGAAATTTTACAGAGAGCCAGATAGCCAAATGGGGTTTATTTTAAATTTGAGCATG 61  
Db 235 AlaGluPheTyrArgGluProAspSerGlnMetGlyValTyrPheLeuAsnLeuSerMet 254  
QY 62 GGGCAGATTTTAAGCTTATTTATGCTAGTCTTTTCATTAGGGATTTTATTGTATGCTACA 121  
Db 255 GlyGlnIleLeuSerLeuPheMetValIleValSerLeuGlyIleLeuLeuTyrAlaThr 274  
QY 122 AAAAATTTCTAAAAATAAAGCAAAATCAA 151  
Db 275 LysAsnSerLysLysIleLysGluAsnGln 284

RESULT 8  
ABM72562  
ID ABM72562 standard; protein; 240 AA.  
XX AC ABM72562;  
XX XX  
XX 20-NOV-2003 (first entry)  
XX Staphylococcus aureus protein #1802.  
XX Antibacterial; vaccine; Gene therapy; infection; sepsis; diagnosis;  
XX enzymatic assay; antibiotic target.  
XX Staphylococcus aureus.  
XX WO200294869-A2.  
XX 28-NOV-2002.  
XX 27-MAR-2002; 2002WO-IB002637.  
XX 27-MAR-2001; 2001GB-00007661.  
XX (CHIR-) CHIRON SPA.  
XX Masignani V, Mora M, Scarselli M;  
XX WPI; 2003-120786/11.  
XX N-PSDB; ACF74122.  
XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
XX preventing Staphylococcal infection, specifically an infection caused by  
XX S. aureus, e.g. sepsis.  
XX Claim 1; SEQ ID NO 3604; 49pp; English.  
XX The invention relates to novel genes and encoded proteins from  
XX Staphylococcus aureus. A composition comprising the S. aureus protein, a  
XX nucleic acid encoding the protein, or an antibody to the protein, is  
XX useful as a pharmaceutical, particularly as a vaccine for treating or  
XX preventing infection due to Staphylococcus bacteria, specifically an  
XX infection caused by S. aureus. The composition is particularly useful for  
XX treating or preventing sepsis in a patient. The composition can also be  
XX used for diagnostics. The protein is also used in an assay for enzymatic  
XX studies and as a target for antibiotics. This sequence represents one of  
XX the novel S. aureus proteins of the invention

SQ Sequence 240 AA;  
Alignment Scores:  
Pred. No.: 3.25e-18 Length: 240  
Score: 234.00 Matches: 60  
Percent Similarity: 49.35% Conservative: 54

```
Best Local Similarity: 25.97% Mismatches: 99
Query Match: 15.01% Indels: 18
DB: 6 Gaps: 6

US-09-647-661-1 (1-900) x ABP72562 (1-240)

QY 136 AATAAAGCAAAATCAATGAATTTTGGATCAGGAAAAAGAGACAACTATTAAACGAG 195
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 AsnLysTyrLysAsnGlyA-GLeuCysSerMetSerAsnMetAsnGlnThrIleMetAsp 27

QY 196 -----GCCATTCTTGAAGATCTTTGACAGCCATTATGAGTTTCTAGTGAA 243
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 28 AlaPheHisPheArgHisAlaThrLysGlnPheAspProGlnLysLysValSerLysGlu 47

QY 244 GAATTAGAAGAAATCGCTGAATCGCCAGGCTATCCCAAGCTTACACACGCGAGCA 303
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 48 AspPheGlnThrIleLeuGluSerGlyArgLeuSerProSerLeuGlyLeuGluPro 67

QY 304 TGCATTCTTGTGTTACTATAAGGATTTAAAAAACAATTCAGTGCACACTAC 363
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 TrpLysPheValIleGlnAspGlnAlaLeuArgAspGluLeuLysAlaHisSerTrp 87

QY 364 TTTAATGAAGAATGATTAAAGCGCTTCAGCGTTAATGGTGTATGCTCTTTAAGACCT 423
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 88 GlyAlaLysGlnLeuAspThrAlaSerHisPheValLeuIlePheAlaArgLys--- 106

QY 424 AGCGAGTTGTATACACACGCGCCATTACATCAAAACCTTTACCCGGAG-----TCTTAT 477
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 ---AsnValThrSerArgSerProTyrValGlnHisMetLeuArgAspIleLysLysTyr 125

QY 478 AAGTTAGAGTATCCCTCTTTTGTCTCAATGCTTGGCGTG-----AGA 522
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 GluAlaGlnThrIleProAlaValGlnGlnLysPheAspAlaPheGlnAlaAspPheHis 145

QY 523 TTCAACACAGCATGCAAGAAATAGAAAGCTATATTATTAGAGCAATGCTATATCGGTGTG 582
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 IleSerAspAsnAspGlnAlaLeuTyrAspTrpSerSerLysGlnThrTyrIleAlaLeu 165

QY 583 GCGCAATTGCGATGGCGGTGAGCTTAATGGGATTCGATTTGCTATTATGAGCTTT 642
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 GlyAsnMetMetThrThrAlaAlaLeuLeuGlyIleAspSerCysProMetGluGlyPhe 185

QY 643 GATCCTTTAAAGTGGTGAAGTTTGAAGAGCGT-----ATCAATAAGCCT 690
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 SerLeuAspThrValThrAspIleLeuAlaAsnLysGlyIleLeuAspThrGluGlnPhe 205

QY 691 AAAATCGATGTTGATCGCTTTGGGCAAGAGGTGCGACAA---CGGAGCCAAAATCA 747
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 GlyLeuSerValMetValAlaPheGlyTyrArgGlnGlnGluProProLysAsnLysThr 225

QY 748 AGAAATCAAAAGCTTGATGCGATTACTTGGTTG 780
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 ArgGlnAlaTyrGluAspValIleGluTrpVal 236

RESULT 9
ABP78332
ID ABP78332 standard; protein; 221 AA.
XX
AC ABP78332;
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae amino acid sequence SEQ ID 3194.
XX
KW Antibacterial; infection; vaccine; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
FN W0200279243-A2.
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB002069.

XX 12-FEB-2001; 2001GB-00003424.
XX (CHIR-) CHIRON SPA.
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX N-PSDB; ABZ39302.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection.
XX Disclosure; Page 422; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention
XX Sequence 221 AA;

Alignment Scores:
Pred No.: 1,85e-15 Length: 221
Score: 210.50 Matches: 58
Percent Similarity: 46.46% Conservative: 47
Best Local Similarity: 25.66% Mismatches: 98
Query Match: 13.50% Indels: 23
DB: 6 Gaps: 4

US-09-647-661-1 (1-900) x ABP78332 (1-221)

QY 151 ATGAAAATTTTGGATTCAGAAAAAGAGACAACTATTAAACGAGCGCCATTCTTGCAG 210
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MetThrValLeuSerLysGluGlnValLeuSerAlaPheLysAsnArgLysSerCysArg 20

QY 211 ATGTTTGACACCCATTATGAGTTTCTAGTGAAGAATTAGAAGAAATCGCTGAATCGCC 270
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 HisTyrAspAlaAlaArgLysIleSerAlaGluAspPheGlnPheIleLeuGluGly 40

QY 271 AGGTATCGCCAACTCTTACACACGAGCATGCGATTTTGTGATGGTTTGTCTTAAAG 330
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 41 ArgLeuSerProSerSerValGlySerGluProTrpGlnPheValValGlnAsnPro 60

QY 331 GATTTAAAAACAATTCGAGTGCAGCTACTTTAATGAAGAAATGATTAAAGCGCT 390
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 GluIleArgGlnAlaIleLysLeuPheSerTrpGlyMetAlaAspAlaLeuAspThrAla 80

QY 391 TCAGCGTTAATGGTGTATGCTCTTTAAGACCTAGCGAGTTGTTACCACGCGCATTAC 450
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 81 SerHisLeuValValPheLeuAlaLysLys-----AsnAlaArgPhe 94

QY 451 ATGCAAAACCTTTACCCGGAGCTTATAAGTTAGA-----GTG 489
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 AspSerProPheMetLeuGluSerLeuLysArgGlyValThrGluProAspAlaVal 114

QY 490 ATCCCTTCTTTTGTCT-----CAATGCTTGGCGTGAAGATTCACACAGC 534
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 GluLysSerLeuAlaArgTyrGlnAlaPheGlnAlaAspAspIleLysIleLeuAspAsp 134

QY 535 ATGCAAGATTAGAAAGCTATATTTAGACAAATGCTATATCGCTGTGGGCAATTTGC 594
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 SerArgAlaLeuPheAspTrpCysArgGlnThrTyrIleAlaLeuAlaAsnMetMet 154

QY 595 ATGGCGGTGAGCTTAATGGGATTCGATGATTGCGATTGAGCGCTTTGATCCCTTTAAA 654
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 ThrGlyAlaAlaMetAlaGlyIleAspSerCysProValGluGlyPheAsnTyrAlaAsp 174

QY 655 GTGGGTGAAGTTTAGAAGAGCGGTATCAAT-----AAGCCTAAATCGCA 699
```

```
Db 175 MetGluArgValLeuSerGlyGlnPheGlyLeuPheAspAlaAlaGluTrpGlyValSer 194
QY 700 TGCTTGATCGTTGGCAAGAGGTGGCAGAACGAGCCAAAATCAAGAAATCAAAA 759
Db 195 ValAlaAlaThrPheGlyTyArgValGlnGlnLeuValThrLysAlaAlaArgProLeu 214
QY 760 GTTGATCGCATTTGG 777
Db 215 GluGluThrValIleTrp 220

RESULT 10
ID AAY90339 standard; protein; 190 AA.
XX AAY90339;
AC AAY90339;
DT 22-NOV-2000 (first entry)
DE H. influenzae nitroreductase.
XX Nitroreductase; CB1954; cytotoxic 4-hydroxylamine; 4HX derivative;
KW monofunctional alkylating agent; prodrug therapy; tumour therapy;
KW non-cytotoxic 2-hydroxylamine derivative; enzyme.
XX Haemophilus influenzae.
OS WO200047725-A1.
PN 17-AUG-2000.
FD 10-FEB-2000; 2000WO-GB0000431.
PF 10-FEB-1999; 99GB-00003019.
PR (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA Minton N, Anlezark G, Vaughan T;
PI WPI; 2000-558214/51.
PT A nitroreductase useful for producing a medicament for anti-tumour therapy
PT preferentially reduces CB1954 to a cytotoxic 4-hydroxylamine derivative
instead of a non-cytotoxic 2-hydroxylamine derivative.
XX Disclosure; Page 44; 56pp; English.
XX This sequence represents a nitroreductase. The invention relates to a
XX nitroreductase characterised in that it preferentially reduces the
XX monofunctional alkylating agent CB1954 to a cytotoxic 4-hydroxylamine
XX (4HX) derivative instead of a non-cytotoxic 2-hydroxylamine derivative.
XX The DNA sequence coding for the nitroreductase can be used in the
XX manufacture of a medicament for prodrug therapy. The nitroreductase,
XX preferably conjugated to a targeting moiety, can be used in the
XX manufacture of a medicament for anti-tumour therapy. The active drug is
XX not administered, it will therefore be less likely to be degraded
XX Sequence 190 AA;

Alignment Scores:
Pred. No.: 9.03e-15 Length: 190
Score: 204.50 Matches: 59
Percent Similarity: 47.89% Conservative: 43
Best Local Similarity: 27.70% Mismatches: 84
Query Match: 13.12% Indels: 27
DB: 3 Gaps: 5

US-09-647-661-1 (1-900) x AAY90339 (1-190)

QY 151 ATGAATTTTGGATCGGAAAAAGAGAACACTATTAAACGAGCGCATTTCTTGAAG 210
Db 1 MetThrGlnLeuThrArgGluGlnValLeuGluLeuPheHisGlnArgSerSerThrArg 20
```

```
QY 211 ATGTTTGACAGCCATTATGAGTTTCTAGTGAAGAATTAGAAGAATCGCTGAATCGCC 270
Db 21 TyrTrpAspProThrLysLysIleSerAspGluAspPheGluCysIleLeuGluCysGly 40
QY 271 AGGCTTATCGCCAAAGCTCTTACAACAGCAGCCATCGCATTTTGTGATGGTTACTAATAG 330
Db 41 ArgLeuSerProSerValGlySerGluProIrrPheLeuValIleGlnAsnLys 60
QY 331 GATTTAAAAAACAATTCGACGTGCACAGTACTTCTTAATGAAGAATGATTAAAGCCCT 390
Db 61 ThrLeuArgGluLysMetLysProPheSerTrpGlyMetIleAsnGlnLeuAspAsnCys 80
QY 391 TCAGCGTTAATGGTGGTATGCTCTTTAAGACCTAGCAGATTGTTACACACGCGCATTTAC 450
Db 81 SerHisLeuValValIleLeuAlaLysLys-----AsnAlaArgTyr 94
QY 451 -----ATGCAAAACCTTTACCCGAGTCTTATAGGTTAGAGTGATCCCTTCTTTGCT 504
Db 95 AspSerGlnGlnAlaAlaLeuThrLysTyrLysAla-----Leu 108
QY 505 CAAATGCTTGGCTGAGATTCAACACAGCATGCAAGATTAGAAAGCTATATTTAGAG 564
Db 109 GlnGluGluAspMetLysLeuLeuGluAsnAspArgThrLeuPheAspTrpCysSerLys 128
QY 565 CAATGCTATATCGTCTGGGCAAAATTTGCATGGCGGTGAGCTTAATCGGATTTGATAGT 624
Db 129 GlnThrTyrIleAlaLeuAlaAsnMetLeuThrGlyAlaSerAlaLeuGlyLeuAspSer 148
QY 625 TGCATTATGGAGGCTTTGATCCCTTTAAAGTGGTGAAGTTTA---GAAGAGCGTATC 681
Db 149 CysProIleGluGlyPheHisTyrAspLysMetAsnGluCysLeuAlaGluGluGlyLeu 168
QY 682 AATAAGCCTMAAATCGCATGCTTGATCGCTTTGGCAAGAGGTTGGCAGAGCGAGCCAA 741
Db 169 PheAspProGln-----GluTyrAlaVal 176
QY 742 AATCAAGAAATCAAAAGTTGATCGGATTACTTGGTTG 780
Db 177 LysSerArgLysGlyLeuAspGluValValLysTrpVal 189

RESULT 11
ID AAY90344 standard; protein; 201 AA.
XX AAY90344;
AC AAY90344;
DT 22-NOV-2000 (first entry)
DE C. jejuni nitroreductase.
XX Nitroreductase; CB1954; cytotoxic 4-hydroxylamine; 4HX derivative;
KW monofunctional alkylating agent; prodrug therapy; tumour therapy;
KW non-cytotoxic 2-hydroxylamine derivative; enzyme.
XX Campylobacter jejuni.
OS WO200047725-A1.
PN 17-AUG-2000.
PD 10-FEB-2000; 2000WO-GB0000431.
PF 10-FEB-1999; 99GB-00003019.
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PI Minton N, Anlezark G, Vaughan T;
PI WPI; 2000-558214/51.
DR N-PSDB; AAA37746.
XX A nitroreductase useful for producing a medicament for anti-tumour therapy
PT preferentially reduces CB1954 to a cytotoxic 4-hydroxylamine derivative
```

[illegible]

```
QY 493 CCTTCTTTTGGTCAAAATGCTTGGGTGAGTATCAACACACAGCATGCAAGATTAGAA--- 549
Db 114 ---LysAlaAlaAsnAspLysGlyArgLysPheSerAlaAspMetHisArgLysAspLeu 132
QY 550 -----AGCTATATTTAGAGCAATGCTATATCGTCTGTGGGCAAAATTTCATG 597
Db 133 HisAspAspAlaGluTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeu 152
QY 598 GGCGTGAGCTTAAGGATTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 645
Db 153 GlyValAlaAlaLeuGlyLeuAspAlaValProIleGluGlyPheAsp 168

RESULT 13
AAW13083
ID AAW13083 standard; protein; 217 AA.
XX AC AAW13083;
XX DT 08-MAY-1997 (first entry)
XX DE Escherichia coli flavin reductase.
XX KW Flavin reductase; nitro reductase; recombinant production; thermostable;
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
XX FT Misc-difference 124 /note= "wild type Phe substituted with Ser"
XX PN JF090000266-A.
XX PD 07-JAN-1997.
XX PF 19-JUN-1995; 95JP-00176736.
XX PR 19-JUN-1995; 95JP-00176736.
XX PA (CHCC ) CHISSO CORP.
XX DR WPI; 1997-112849/11.
XX DR N-PSDB; AAT61841.
XX PT Escherichia coli flavin reductase - prepd. by mutating wild type E. coli
XX PS Claim 7; Page 7; 9pp; Japanese.
XX CC The present sequence is an Escherichia coli flavin reductase (FR), which
XX CC was prepared by mutating an E. coli derived wild type nitro reductase
XX CC (NR) gene. The FR gene can be used for the recombinant production of the
XX CC new stable, high activity FR. The recombinant plasmid pNR1DNA was used as
XX CC the template for the random mutation of the E. coli NR gene by PCR. E.
XX CC coli containing a plasmid capable of expressing the mutated NR gene were
XX CC prepared, and an E. coli containing a plasmid expressing the FR gene
XX CC selected. The FR encoded by pNR247 was purified, found to be stable at
XX CC temperatures of up to 40 degrees C and the base sequence of its gene
XX CC determined
XX SQ Sequence 217 AA;

Alignment Scores:
Pred. No.: 2,45e-13 Length: 217
Score: 192.50 Matches: 49
Percent Similarity: 48.30% Conservative: 36
Best Local Similarity: 27.84% Mismatches: 50
Query Match: 12.35% Indels: 41
DB: 2 Gaps: 5

US-09-647-661-1 (1-900) x AAW13083 (1-217)
QY 193 GAGCGCCATCTTGCACAGATGTTGACGCCATATGATGATTTCTAGTGAGAAATTAGAA 252
```

```
Db 9 LysArgHisSerThrLysAlaPheAspAlaSerLysLysLeuThrProGluGlnAlaGlu 28
QY 253 GAATCGCTGAAATCGCCAGGCTATCGCAAGCTCTTACACACGACGCGCATGCGCATTTT 312
Db 29 GlnIleLysThrLeuGlnTyrSerProSerSerThrAsnSerGlnProTrpHisPhe 48
QY 313 GTGATGGTTACTAATAAGGATTTAAAAAACAATTCAGTGCACAGC----- 360
Db 49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnTyr 68
QY 361 TACTTTAATGAAGAAATGATTAAAGCGCTTCACGCTTAATGGTGGTATGCTCTTTAAGA 420
Db 69 ValPheAsnGluArgLysMetLeuAspAlaSerHisValValValPheCysAlaLysThr 88
QY 421 CCTAGCGAGTGTGTTACCACACAGCGCATTTACATGCAAAACCTTTTAC----- 465
Db 89 Ala-----MetAspAspValTyrPheLysLeuValVal 99
QY 466 -----CCGAGTCTTATAAGTTAGGTAGTGCATC 492
Db 100 AspGlnGluAspAlaAspGlyArgPheAlaThrProGluAla----- 113
QY 493 CCTCTTTTGTCTCAAAATGCTTGGCGTGAGATTCAACACACAGCATGCAAGATTAGAA--- 549
Db 114 --LysAlaAlaAsnAspLysGlyArgLysPheSerAlaAspMetHisArgLysAspLeu 132
QY 550 -----AGCTATATTTAGCAATGCTATATCGTCTGTGGGCAAAATTTCATG 597
Db 133 HisAspAspAlaGluTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeu 152
QY 598 GGCGTGAGCTTAATGCGGATTGGATGATGATGATGATGATGATGATGATGATGATGAT 645
Db 153 GlyValAlaAlaLeuGlyLeuAspAlaValProIleGluGlyPheAsp 168

RESULT 14
ABP70750
ID ABP70750 standard; protein; 217 AA.
XX AC ABP70750;
XX DT 15-MAY-2003 (first entry)
XX DE Mutant E.coli nitroreductase, NFSB, Y68G/F124Q.
XX KW Nitroreductase; enzyme; cytostatic; NFSB; dihydropteridine reductase;
XX KW oxygen-insensitive NAD(P)H nitroreductase; NTR; NFSB; prodrug CD1954;
XX KW anti-tumour; cancer; mutant; mutain.
XX OS Escherichia coli.
XX FH Synthetic.
XX FT Key Location/Qualifiers
XX FT Misc-difference 68
XX FT Misc-difference 124 /note= "Wild-type Tyr substituted by Gly"
XX FT Misc-difference 124 /note= "Wild-type Phe substituted by Gln"
XX PN WO2003018788-A2.
XX PD 06-MAR-2003.
XX PF 21-AUG-2002; 2002WO-GB003833.
XX PR 21-AUG-2001; 2001GB-00020294.
XX PR 06-SEP-2001; 2001GB-00021662.
XX PR 03-OCT-2001; 2001US-0326846P.
XX PA (MLML-) ML LAB PLC.
XX PI Grove JI, Searle PF, Lovering AL;
XX DR WPI; 2003-300732/29.
```



XX New recombinant mutant nitroreductase having increased nitroreductase  
PT activity, useful as a medicament for treating cancer, for converting a  
PT prodrug into a cytotoxic agent, and in designing or screening for  
PT improved prodrugs.  
XX PS Claim 15; Page; 63pp; English.  
XX  
CC The present invention relates to mutant nitroreductases (ABP70750-  
CC ABP70759), which have increased nitroreductase activity as compared to  
CC the wild-type enzyme, E. coli NFSB nitroreductase (EC 1.6.99.7, also  
CC known as oxygen-insensitive NAD(P)H nitroreductase or dihydropteridine  
CC reductase (NTR)). NFSB reduces the aziridinyl prodrug CD1954 within cells  
CC to produce a genotoxic 4-hydroxylamine derivative and so is useful as an  
CC anti-tumour agent. The mutant nitroreductases are therefore useful as  
CC medicaments for treating cancer. Note: The present sequence is not shown  
CC in the specification but is derived from the Escherichia coli wild-type  
CC nitroreductase sequence given in figure 9 (see ABP70750)

XX SQ Sequence 217 AA;

Alignment Scores:  
Pred. No.: 2,45e-13 Length: 217  
Score: 192.50 Matches: 50  
Percent Similarity: 47.16% Conservative: 33  
Best Local Similarity: 28.41% Mismatches: 52  
Query Match: 12.35% Indels: 41  
DB: 6 Gaps: 5

US-09-647-661-1 (1-900) x ABP70750 (1-217)

QY 193 GAGCGCCATTCTTCACAGATGTTTGACGCCATTATGAGTTTCTAGTGAAGAATTAGAA 252  
Dy 9 LysArgHisSerThrLysAlaPheAspAlaSerLysLysLeuThrProGluGlnAlaGlu 28  
QY 253 GAATTCGCTGAATCGCGAGCTATCGCAAGCTTTACACAGCGCCATGCGCATTTT 312  
Dy 29 GlnIleLysThrLeuLeuGlnTyrSerProSerSerThrAsnSerGlnProTrpHisPhe 48  
QY 313 GTGATGTTTACTAATAGGATTTAA--AAACAATTCAGTGCACAGC 360  
Dy 49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnGly 68  
QY 361 TACTTTAATGAAGAATGATTAAAGCGCTTCAGCGTTATGTTGTTGCTCTTTAAGA 420  
Dy 69 ValPheAsnGluArgLysMetLeuAspAlaSerHisValValPheCysAlaLysThr 88  
QY 421 CCTAGCGAGTTGTTACCACAGCGCCATTACATGCAAAACCTTTAC----- 465  
Dy 89 Ala-----MetAspAspValTrpLeuLysLeuValVal 99  
QY 466 -----CCGAGTCTTATAGGTTAGTGATC 492  
Dy 100 AspGlnGluAspAlaAspGlyArgPheAlaThrProGluAla----- 113  
QY 493 CTTCTTTTGTCAATGCTTGGCGTGAGATTCAACACAGCATGCAAGATTAGAA--- 549  
Dy 114 ---LysAlaAlaAsnAspLysGlyArgLysPheGlnAlaAspMethHisArgLysAspLeu 132  
QY 550 -----ACGTATATTTAGACAAATGCTATATCCGTGGGGGCAAAATTCGATG 597  
Dy 133 HisAspAspAlaGluTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeu 152  
QY 598 GCGGTGAGCTTAATGGGATGATAGTTGCAATTATGGAGCTTTGAT 645  
Dy 153 GlyValAlaAlaLeuGlyLeuAspAlaValProIleGluGlyPheAsp 168

RESULT 15

ABP70759

ID ABP70759 standard; protein; 217 AA.

XX

AC

XX

DT 15-MAY-2003 (first entry)  
XX  
DE Mutant E.coli nitroreductase, NFSB, #7.  
XX  
KW Nitroreductase; enzyme; cytostatic; NFSB; dihydropteridine reductase;  
KW oxygen-insensitive NAD(P)H nitroreductase; NTR; NFSB; prodrug CD1954;  
KW anti-tumour; cancer; mutant; mutein.  
XX  
OS Escherichia coli.  
OS Synthetic.  
FH  
FH Key Location/Qualifiers  
FT Misc-difference 124  
FT /note= "Wild-type Phe can be substituted by Asn, Cys,  
FT Gly, Lys, Met, Trp, Tyr, Ala, Gln, His, Ile, Leu, Ser,  
FT Thr or Val"  
XX  
PN WO2003018788-A2.  
XX  
PD 06-MAR-2003.  
XX  
XX 21-AUG-2002; 2002WO-GB0038333.  
XX  
PR 21-AUG-2001; 2001GB-00020294.  
PR 06-SEP-2001; 2001GB-00021662.  
PR 03-OCT-2001; 2001US-0326846P.  
XX  
XX (MLML-) ML LAB PLC.  
XX  
XX Grove JT, Searle PF, Lovering AJ;  
XX WPI; 2003-300732/29.  
XX  
PT New recombinant mutant nitroreductase having increased nitroreductase  
PT activity, useful as a medicament for treating cancer, for converting a  
PT prodrug into a cytotoxic agent, and in designing or screening for  
PT improved prodrugs.  
XX  
XX Claim 23; Page; 63pp; English.  
PS  
CC The present invention relates to mutant nitroreductases (ABP70750-  
CC ABP70759), which have increased nitroreductase activity as compared to  
CC the wild-type enzyme, E. coli NFSB nitroreductase (EC 1.6.99.7, also  
CC known as oxygen-insensitive NAD(P)H nitroreductase or dihydropteridine  
CC reductase (NTR)). NFSB reduces the aziridinyl prodrug CD1954 within cells  
CC to produce a genotoxic 4-hydroxylamine derivative and so is useful as an  
CC anti-tumour agent. The mutant nitroreductases are therefore useful as  
CC medicaments for treating cancer. Note: The present sequence is not shown  
CC in the specification but is derived from the Escherichia coli wild-type  
CC nitroreductase sequence given in figure 9 (see ABP70750)

XX SQ Sequence 217 AA;

Alignment Scores:  
Pred. No.: 4,21e-13 Length: 217  
Score: 190.50 Matches: 49  
Percent Similarity: 47.73% Conservative: 35  
Best Local Similarity: 27.84% Mismatches: 51  
Query Match: 12.22% Indels: 41  
DB: 6 Gaps: 5

US-09-647-661-1 (1-900) x ABP70759 (1-217)

QY 193 GAGCGCCATTCTTCACAGATGTTTGACGCCATTATGAGTTTCTAGTGAAGAATTAGAA 252  
Dy 9 LysArgHisSerThrLysAlaPheAspAlaSerLysLysLeuThrProGluGlnAlaGlu 28  
QY 253 GAATTCGCTGAATCGCGAGCTATGCGCAAGCTTTACACAGCGCCATGCGCATTTT 312  
Dy 29 GlnIleLysThrLeuLeuGlnTyrSerProSerSerThrAsnSerGlnProTrpHisPhe 48  
QY 313 GTGATGTTTACTAATAGGATTTAA-----CAAAATTCGAGTGCACAGC----- 360

Db 49 IleValAlaSerThrGluGluGlyLysAlaArgValAlaLysSerAlaAlaGlyAsnTyr 68  
QY 361 TACTTTAATGAAGAATGATTAARAGCGCTTCAGCGTTAATGCTGGTATGCTCTTTAAGA 420  
Db 69 ValPheAsnGluArgGlyMetLeuAspAlaSerHisValValPheCysAlaLysThr 88  
QY 421 CCTACCGAGTTGTTACCAACGCGCATTCACATGCAAAACCTTTAC----- 465  
Db 89 Ala-----MetAspAspValTrpLeuLysLeuValVal 99  
QY 466 -----CCGGAGTCTTATAGGTTAGAGTCATC 492  
Db 100 AspGlnGluAspAlaAspGlyArgPheAlaThrProGluAla----- 113  
QY 493 CTTCTTTTGTCAATGCTTGGCGTGAGATTCAACACAGCATGCAAGAGATTAGAA--- 549  
Db 114 ---LysAlaAlaAsnAspLysGlyArgLysPhe\*\*\*AlaAspMetHisArgLysAspLeu 132  
QY 550 -----AGCTATATTTAGAGCAATGCTATATCGCTGTGGGCAAAATTTGCATG 597  
Db 133 HisAspAspAlaGluTrpMetAlaLysGlnValTyrLeuAsnValGlyAsnPheLeuLeu 152  
QY 598 GCGGTGAGCTTAATGGGATTGGATAGTTGCATTATTGGAGGCTTTGAT 645  
Db 153 GlyValAlaAlaLeuGlyLeuAspAlaValProIleGluGlyPheAsp 168

Search completed: April 20, 2004, 19:24:49  
Job time : 64 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 20, 2004, 19:11:02 ; Search time 64 Seconds  
(without alignments)  
7753.307 Million cell updates/sec

Title: US-09-647-661-1

Perfect score: 1559

Sequence: 1 tgcagaattttacagagagc.....ggggtgtttcaagcggttc 900

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1124875 seqs, 275673149 residues

Total number of hits satisfying chosen parameters: 2249750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DB=US09647661/runat\_20042004\_135346\_2594/app\_query.fasta\_1.1095  
-O=/cgn2\_1/USPTO\_spool\_p/US09647661/runat\_20042004\_135346\_2594/app\_query.fasta\_1.1095  
-DB=Published Applications AA -GFM=fastan -SUFFIX=xapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -SPART=1 -END=-1 -MATRIX=blomax62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct.THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MOD=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09647661@cgn\_1.13 @runat\_20042004\_135346\_2594  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/FCI\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/ACTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description

RESULT 1

US-09-882-227-180

Sequence 180, Application US/09882227

Publication No. US20030158396A1

GENERAL INFORMATION:

APPLICANT: Kleantous, Harold

APPLICANT: Al-Garawi, Amal

APPLICANT: Miller, Charles

APPLICANT: Tomb, Jean-Francois

APPLICANT: Oomen, Raymond P

TITLE OF INVENTION: Identification of Polynucleotides

TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in th

TITLE OF INVENTION: Genome

FILE REFERENCE: 06132/047002

CURRENT APPLICATION NUMBER: US/09/882,227

CURRENT FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 08/902,615

PRIOR FILING DATE: 1997-07-29

NUMBER OF SEQ ID NOS: 638

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 180

LENGTH: 210

TYPE: PRT

ORGANISM: Helicobacter pylori

08/19021 6/15/1997

## US-09-882-227-180

## Alignment Scores:

Pred. No.: 1.59e-110 Length: 210  
Score: 1072.00 Matches: 207  
Percent Similarity: 99.05% Conservatives: 1  
Best Local Similarity: 98.57% Mismatches: 2  
Query Match: 68.78% Indels: 0  
DB: 10 Gaps: 0

US-09-647-661-1 (1-900) x US-09-882-227-180 (1-210)

QY 151 ATGAAATTTTGGATCAGGAAAAAGACAACTATTAAACGAGCGCCATTCTTGCAG 210  
DB 1 MetLysPheLeuAspGlnGlyArgGlnLeuLeuAsnGluArgHisSerCysLys 20  
QY 211 ATGTTTGACGCCATTATGAGTTTCTAGTGAAGAAATTAGAGAAATCGCTCAATCGCC 270  
DB 21 MetPheAspSerHisTyrGluPheSerThrGluLeuGluGluAlaGluAla 40  
QY 271 AGGCTATGCCCAAGCTCTTACACACGAGCGCATGGCATTGTGATGGTTACTAATAAG 330  
DB 41 ArgLeuSerProSerSerTyrAsnThrGlnProThrHisPheValMetValThrAspLys 60  
QY 331 GATTTAAAAACAAATTCAGTGCACAGCTACTTTAATGAAGAAATGATTAAGAGCGCT 390  
DB 61 AspLeuLysLysGlnIleAlaAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80  
QY 391 TCAGCGTTAATGGTGTATGCTTTAAGACTAGCGAGTGTGTACACACGCGCATTAC 450  
DB 81 SerAlaLeuMetValCysSerLeuArgProSerGluLeuLeuProHisGlyHisTyr 100  
QY 451 ATGCAAAACCTTTACCCGGAGTCTTATAAGGTAGAGTATCCCTCTTTTGCCTCAATG 510  
DB 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120  
QY 511 CTGGCGTCAGATTCAACACAGCATGCAAGATTAGAAAGCTATATTAGACCAATGC 570  
DB 121 LeuGlyValArgPheAsnHisSerMetGlnArgLeuGluSerTyrIleLeuGluGlnCys 140  
QY 571 TATATCGCTGTGGGCAAAATTTGCATGGCGGTGAGCTTAATGGGATTGGATTTGCATT 630  
DB 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
QY 631 ATGAGGCTTTGATCCCTTTAAAGTGGTGAAGTTTAGAGAGCGTATCAATAAGCCT 690  
DB 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro 180  
QY 691 AAAATCGCATGCTTGCATCGCTTTGGCAAGAGGGTGGCAGAGCGAGCAAAATCAAGA 750  
DB 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
QY 751 AAATCAAAAGTTGATCGGATTACTTGGTTG 780  
DB 201 LysSerLysValAspAlaIleThrTrpLeu 210

## RESULT 2

US-10-364-397-29  
; Sequence 29, Application US/10364397  
; Publication No. US20040014191A1  
; GENERAL INFORMATION:  
; APPLICANT: Minton, Nigel  
; APPLICANT: Anlezark, Gill  
; APPLICANT: Vaughan, Thomas  
; TITLE OF INVENTION: Nitroreductase Enzymes  
; FILE REFERENCE: 1581.0350001  
; CURRENT APPLICATION NUMBER: US/10/364,397  
; PRIORITY FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: 09/913,068  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: PCT/GB00/00431  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: UK 9903019.9

; PRIOR FILING DATE: 1999-02-10  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 29  
; LENGTH: 210  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-10-364-397-29

## Alignment Scores:

Pred. No.: 5.73e-110 Length: 210  
Score: 1067.00 Matches: 206  
Percent Similarity: 99.05% Conservatives: 2  
Best Local Similarity: 98.10% Mismatches: 2  
Query Match: 68.44% Indels: 0  
DB: 15 Gaps: 0

US-09-647-661-1 (1-900) x US-10-364-397-29 (1-210)

QY 151 ATGAAATTTTGGATCAGGAAAAAGACAACTATTAAACGAGCGCCATTCTTGCAG 210  
DB 1 MetLysPheLeuAspGlnGlyArgGlnLeuLeuAsnGluArgHisSerCysLys 20  
QY 211 ATGTTTGACGCCATTATGAGTTTCTAGTGAAGAAATTAGAGAAATCGCTCAATCGCC 270  
DB 21 MetPheAspSerHisTyrGluPheSerThrGluLeuGluGluAlaGluAla 40  
QY 271 AGGCTATGCCCAAGCTCTTACACACGAGCGCATGGCATTGTGATGGTTACTAATAAG 330  
DB 41 ArgLeuSerProSerSerTyrAsnThrGlnProThrHisPheValMetValThrAspLys 60  
QY 331 GATTTAAAAACAAATTCAGTGCACAGCTACTTTAATGAAGAAATGATTAAGAGCGCT 390  
DB 61 AspLeuLysLysGlnIleAlaAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80  
QY 391 TCAGCGTTAATGGTGTATGCTTTAAGACTAGCGAGTGTGTACACACGCGCATTAC 450  
DB 81 SerAlaLeuMetValCysSerLeuLysProSerGluLeuLeuProHisGlyHisTyr 100  
QY 451 ATGCAAAACCTTTACCCGGAGTCTTATAAGGTAGAGTATCCCTCTTTTGCCTCAATG 510  
DB 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120  
QY 511 CTGGCGTCAGATTCAACACAGCATGCAAGATTAGAAAGCTATATTAGACCAATGC 570  
DB 121 LeuGlyValArgPheAsnHisSerMetGlnLysLeuGluSerTyrIleLeuGluGlnCys 140  
QY 571 TATATCGCTGTGGGCAAAATTTGCATGGCGGTGAGCTTAATGGGATTGGATTTGCATT 630  
DB 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
QY 631 ATGAGGCTTTGATCCCTTTAAAGTGGTGAAGTTTAGAGAGCGTATCAATAAGCCT 690  
DB 161 IleGlyGlyPheAspProLeuLysValGlyGluValLeuGluGluArgIleAsnLysPro 180  
QY 691 AAAATCGCATGCTTGCATCGCTTTGGCAAGAGGGTGGCAGAGCGAGCAAAATCAAGA 750  
DB 181 LysIleAlaCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerGlnLysSerArg 200  
QY 751 AAATCAAAAGTTGATCGGATTACTTGGTTG 780  
DB 201 LysSerLysValAspAlaIleThrTrpLeu 210

## RESULT 3

US-10-335-977-6487  
; Sequence 6487, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:  
 ADDRESS: LAHIVE & COCKFIELD  
 STREET: 28 State Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: Windows NT 4.0  
 SOFTWARE: UNIX  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/335,977  
 FILING DATE: 30-Dec-2002  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/993,002  
 FILING DATE: 17-DEC-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mandragoras, Amy E.  
 REGISTRATION NUMBER: 36,207  
 REFERENCE/DOCKET NUMBER: GTN-018  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 742-4214  
 INFORMATION FOR SEQ ID NO: 6487:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 210 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Helicobacter pylori  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...210  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6487:

Alignment Scores:  
 Pred. No.: 4,58e-107 Length: 210  
 Score: 1041.00 Matches: 200  
 Percent Similarity: 97.14% Conservative: 4  
 Best Local Similarity: 95.24% Mismatches: 6  
 Query Match: 66.77% Indels: 0  
 DB: 12 Gaps: 0

US-09-647-661-1 (1-900) x US-10-335-977-6487 (1-210)

QY 151 ATGAATTTTGGATCAGGAAAAAGACACAACTATTAAACGAGCGCCATTCTTGCAAG 210  
 DB 1 MetLysPheLeuAspHisGluLysArgGlnLeuLeuAsnGluArgHisSerCysLys 20  
 QY 211 ATGTTTGACGCAATTATGAGTTTCTAGTGAAGATTAGAAGAAATCGCTGAATCGCC 270  
 DB 21 MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluGluLeuAlaGluLeuAla 40  
 QY 271 AGGCTATCGCAAGCTTTACACACGACGACGATTTGTGATGTTACTAATAAG 330  
 DB 41 ArgLeuSerProSerSerTyrAsnThrGlnProThrPhePheValMetValThrAsnLys 60  
 QY 331 GATTTAAAAAACAATTCAGTCGACGCTACTTTTATGAAGAAATGATTTAAAGCGCT 390  
 DB 61 AspLeuLysLysGlnIleAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80  
 QY 391 TCAGCGTTAATGGTGTATGCTCTTTAAGACCTAGCGAGTTGTTCACACGCGCATAC 450  
 DB 81 SerAlaLeuMetValValCysSerLeuLysProSerGluLeuLeuProThrSerHisTyr 100  
 QY 451 ATGCAAAACCTTTACCGGAGCTTATAGGTTAGTAGTATCCCTTTCTTTCTCAATG 510  
 DB 101 MetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMet 120

QY 511 CTTGGCGTGAGATTCAACACAGCATGCAAGATTAGAAAGCTATATTTAGAGCAATGC 570  
 DB 121 LeuGlyValArgPheAsnHisSerMetGlnLysLeuGluSerTyrIleLeuGluGlnCys 140  
 QY 571 TATATCGCTGTGGGCAAAATTTGCATGGCGTGAGCTTAATGGGATTGGATAGTTCGATT 630  
 DB 141 TyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIle 160  
 QY 631 ATTGGAGCGCTTTCATCCTTTAAAGTGGGTGAAGTTTATAGAAGCGGTATCAATAAGCCT 690  
 DB 161 IleGlyGlyPheAspProLeuLysValGlyGluIleLeuGluGluArgIleAsnLysPro 180  
 QY 691 AAAATCGCATGCTTGTATCGCTTTGGCAAGAGGCTGGCAGACGAGCCAAATAATCAAGA 750  
 DB 181 LysIleValCysLeuIleAlaLeuGlyLysArgValAlaGluAlaSerLysLysSerArg 200  
 QY 751 AAATCAAAAGTTCATGCGATTACTTCGTTG 780  
 DB 201 LysSerLysValAspAlaIleThrTrpLeu 210

# RESULT 4

US-10-335-977-6486  
 ; Sequence 6486, Application US/10335977  
 ; Publication No. US20040052799A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DOUGLAS SMITH et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
 ; RELATING TO HELICOBACTER PYLORI FOR  
 ; DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 10031  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: LAHIVE & COCKFIELD  
 ; STREET: 28 State Street  
 ; CITY: Boston  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02109-1875  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: Windows NT 4.0  
 ; SOFTWARE: UNIX  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/335,977  
 ; FILING DATE: 30-Dec-2002  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/993,002  
 ; FILING DATE: 17-DEC-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mandragoras, Amy E.  
 ; REGISTRATION NUMBER: 36,207  
 ; REFERENCE/DOCKET NUMBER: GTN-018  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 227-7400  
 ; TELEFAX: (617) 742-4214  
 ; INFORMATION FOR SEQ ID NO: 6486:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 134 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: YES  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Helicobacter pylori  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (B) LOCATION 1...134  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6486:

US-10-335-977-6486

Alignment Scores:  
 Pred. No.: 6.47e-62 Length: 134

Score: 636.00 Matches: 122  
Percent Similarity: 95.52% Conservative: 6  
Best Local Similarity: 91.04% Mismatches: 6  
Query Match: 40.80% Indels: 0  
DB: 12 Gaps: 0

US-09-647-661-1 (1-900) x US-10-335-977-6486 (1-134)

QY 379 ATTAAAGCGCTTACGGTAAATGGTGTATGCTTTTAAGACCTAGCGAGTTGTACCA 438  
: : : : :  
Db 1 LeuYsAlaLeuGlnLeuLeuMetValValCysSerLeuYsProSerGluLeuLeuPro 20  
QY 439 CACGGGCATTACATGCAAAACCTTTACCGGAGCTTATAGGTTAGAGTGCCTTCT 498  
: : : : :  
Db 21 ThrSerHisTyrMetGlnAsnLeuTyrProGluSerTyrLysValArgValIleProSer 40  
QY 499 TTTGCTCAATGCTTGGCGTGAAGTCAACACGAGCATGCAAAAGATTAGAAAGCTATAT 558  
: : : : :  
Db 41 PheAlaGlnMetLeuGlyValArgPheAsnHisSerMetGlnLysLeuGluSerTyrIle 60  
QY 559 TTAGACATGCTATATCGCTGTGGGCAAAATTTGATGGGCTGAGCTTATGGGATG 618  
: : : : :  
Db 61 LeuGluGlnCysTyrIleAlaValGlyGlnIleCysMetGlyValSerLeuMetGlyLeu 80  
QY 619 GATAGTTGCATTATTGGAGGCTTTGATCCTTTAAAGTGGTCAAGTTTGAAGAGCGT 678  
: : : : :  
Db 81 AspSerCysIleIleGlyGlyPheAspProLeuLysValGlyGluIleLeuGluArg 100  
QY 679 ATCAATAGCTAAATCGCATGCTTGCATGCTTGGCAAGAGGTGCGAAGCGAGC 738  
: : : : :  
Db 101 IleAsnLysProLysIleValCysLeuIleAlaLeuGlyLysArgValAlaGluLysSer 120  
QY 739 CAAAATCAAGAAATCAAAAGTTGATCGGATTACTTGGTTG 780  
: : : : :  
Db 121 LysLysSerArgLysSerLysValAspAlaIleThrIleLeu 134

RESULT 5

US-10-335-977-6485  
; Sequence 6485, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/993,002  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragoras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: GTN-018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 742-4214  
; INFORMATION FOR SEQ ID NO: 6485:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 127 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...127  
SEQUENCE DESCRIPTION: SEQ ID NO: 6485:  
US-10-335-977-6485

Alignment Scores:  
Pred. No.: 8,33e-61 Length: 127  
Score: 626.00 Matches: 120  
Percent Similarity: 97.64% Conservative: 4  
Best Local Similarity: 94.49% Mismatches: 3  
Query Match: 40.15% Indels: 0  
DB: 12 Gaps: 0

US-09-647-661-1 (1-900) x US-10-335-977-6485 (1-127)

QY 400 ATGTGTATGCTCTTTTAAGACCTAGCGAGTTGTACCACACGCGCCATTACATGCAAAAC 459  
: : : : :  
Db 1 MetValValCysSerLeuLysProSerGluLeuLeuProThrSerHisTyrMetGlnAsn 20  
QY 460 CTTTACCGGAGCTTATAAGGTTAGAGTATCCCTCTTTTGTCTCAAATGCTTGGCGTG 519  
: : : : :  
Db 21 LeuTyrProGluSerTyrLysValArgValIleProSerPheAlaGlnMetLeuGlyVal 40  
QY 520 AGATTCAACACACGAGCATGCAAAAGATTAGAAAGCTATATTTAGAGCAATGCTATATCGCT 579  
: : : : :  
Db 41 ArgPheAsnHisSerMetGlnLysLeuGluSerTyrIleLeuGluGlnCysTyrIleAla 60  
QY 580 GTGGGGCAAAATTTGCATGGCGGCTGAGCTTAAATGGGATTGGATGTCATATTGAGGC 639  
: : : : :  
Db 61 ValGlyGlnIleCysMetGlyValSerLeuMetGlyLeuAspSerCysIleIleGlyGly 80  
QY 640 TTTGATCCTTTAAAGTGGGTGAAGTTTAAAGAGCGGTATCAATAAGCCTAAATCGCA 699  
: : : : :  
Db 81 PheAspProLeuLysValGlyGluIleLeuGluArgIleAsnLysProLysIleVal 100  
QY 700 TGCTTGATCGCTTTGGCAAGAGGTGGCAGAGCGAGCCGCAAAATCAAGAAATCAAAA 759  
: : : : :  
Db 101 CysLeuIleAlaLeuGlyLysArgValAlaGluLysSerLysLysSerArgLysSerLys 120  
QY 760 GTTGATCGGATTACTTGGTTG 780  
: : : : :  
Db 121 ValAspAlaIleThrIleLeu 127

RESULT 6

US-10-335-977-6484  
; Sequence 6484, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX

COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 6483:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 91 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...91  
SEQUENCE DESCRIPTION: SEQ ID NO: 6483:  
US-10-335-977-6483  
Alignment Scores:  
Pred. No.: 1,23e-38 Length: 91  
Score: 427.00 Matches: 86  
Percent Similarity: 95.60% Conservative: 1  
Best Local Similarity: 94.51% Mismatches: 4  
Query Match: 27.39% Indels: 1  
DB: 12 Gaps: 0  
US-09-647-661-1 (1-900) x US-10-335-977-6483 (1-91)  
QY 151 ATGAAATTTTGGATCAGGAAAGAAAGAACAACTATTAAACGAGCGCCATTCTTCCAAG 210  
Db 1 MetLysPheLeuAspHisGluGluA-gA-gGlnLeuLeuAsnGluA-gHisSerCysLys 20  
QY 211 ATGTTTGACAGCCATTATGAGTTTCTAGTGAAGATTAGAACAAATCGCTGAATCGCC 270  
Db 21 MetPheAspSerHisTyrGluPheSerSerThrGluLeuGluGluLeuAlaGluIleAla 40  
QY 271 AGGCTATCGCAAGCTCTTACAAACAGCGCCATGCGCATTTTGTGATGGTTACTAATAAG 330  
Db 41 ArgLeuSerProSerSerTyrAsnThrGlnProTyrHisPheValMetValThrAsnLys 60  
QY 331 GATTTAAAAACAAATTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 390  
Db 61 AspLeuLysLysGlnIleAlaAlaHisSerTyrPheAsnGluGluMetIleLysSerAla 80  
QY 391 TCAGC-GTTAATGGTGTGCTGCTCTTTAAGACC 422  
Db 81 SerAlaValAsnGlyGlyMetLeuPheLysThr 91  
RESULT 8  
US-09-895-913A-102  
Sequence 102, Application US/09895913A  
Patent No. US20020160456A1  
GENERAL INFORMATION:  
APPLICANT: Kleanthous, Harold  
APPLICANT: Al-Garawi, Amal  
APPLICANT: Miller, Charles

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 6484:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 97 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...97  
SEQUENCE DESCRIPTION: SEQ ID NO: 6484:  
US-10-335-977-6484  
Alignment Scores:  
Pred. No.: 5.69e-41 Length: 97  
Score: 448.00 Matches: 90  
Percent Similarity: 94.85% Conservative: 2  
Best Local Similarity: 92.78% Mismatches: 5  
Query Match: 28.74% Indels: 1  
DB: 12 Gaps: 0  
US-09-647-661-1 (1-900) x US-10-335-977-6484 (1-97)  
QY 133 AAAAAATAAGGAAATCAATGAAATTTTGGATCAGGAAAGAAAGAACAACTATTAAAC 192  
Db 1 LysAsnLysGlyGluIleMetLysPheLeuAspHisGluGluA-gA-gGlnLeuLeuAsn 20  
QY 193 GAGCGCCATCTTCAAGATGTTTCACAGCCATTATGAGTTTCTAGTGAAGATTAGAA 252  
Db 21 GluArgHisSerCysLysMetPheAspSerHisTyrGluPheSerThrGluLeuGlu 40  
QY 253 GAAATCGCTGAATCGCAGGCTATCGCAAGCTCTTACAAACAGCGCCATGCGCATTTT 312  
Db 41 GluIleAlaGluIleAlaArgLeuSerProSerSerTyrAsnThrGlnProTyrHisPhe 60  
QY 313 GTGATGGTTACTAATAAGATTAAAAACAAATTCAGTCAGTCAGTCAGTCAGTCAGTCAG 372  
Db 61 ValMetValThrAsnLysAspLeuLysLysGlnIleAlaAlaHisSerTyrPheAsnGlu 80  
QY 373 GAAATGATTAAAGCGCTTCAGC-GTTAATGGTGTGCTGCTCTTTAAGACC 422  
Db 81 GluMetIleLysSerAlaSerAlaValAsnGlyGlyMetLeuPheLysThr 97  
RESULT 7  
US-10-335-977-6483  
Sequence 6483, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESS: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts

APPLICANT: Tomb, Jean Francois  
APPLICANT: Comen, Raymond P.  
TITLE OF INVENTION: Identification of Polynucleotides  
TITLE OF INVENTION: Encoding No. US20020160454a1 Helicobacter Polypeptides in the  
TITLE OF INVENTION: Genome  
FILE REFERENCE: 06132/043002  
CURRENT APPLICATION NUMBER: US/09/895,913A  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 08/891,227  
PRIOR FILING DATE: 1997-06-24  
NUMBER OF SEQ ID NOS: 368  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 102  
LENGTH: 284  
TYPE: PRT  
ORGANISM: Helicobacter pylori  
US-09-895-913A-102

Alignment Scores:  
Pred. No.: 2,76e-18 Length: 284  
Score: 246.00 Matches: 49  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.00% Mismatches: 0  
Query Match: 15.78% Indels: 0  
DB: 9 Gaps: 0

US-09-647-661-1 (1-900) x US-09-895-913A-102 (1-284)

QY 2 GCAGATTTTACAGAGCCAGATAGCCAAATGGGGTTTATTTTAAATTTGAGCATG 61  
Db 235 AlaGluPheTyArgGluProAspSerGlnMetGlyValTyPheLeuAsnLeuSerMet 254  
QY 62 GGCAGATTTTAACTTTATTTAGTGTGTTTATTTAGGATTTTATGATGCTACA 121  
Db 255 GlyGlnIleLeuSerLeuPheMetValIleValSerLeuGlyIleLeuLeuTyzAlaThr 274  
QY 122 AAAAAATCTAAAAAATAAGGAAATCAA 151  
Db 275 LysAsnSerLysLysIleLysGluAsnGln 284

RESULT 9  
US-10-335-977-5526  
Sequence 5526, Application US/10335977  
Publication No. US20040052799A1  
GENERAL INFORMATION:  
APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 5526:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 283 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...283  
SEQUENCE DESCRIPTION: SEQ ID NO: 5526:  
US-10-335-977-5526

Alignment Scores:  
Pred. No.: 2.18e-16 Length: 283  
Score: 229.00 Matches: 45  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 93.75% Mismatches: 0  
Query Match: 14.69% Indels: 0  
DB: 12 Gaps: 0

US-09-647-661-1 (1-900) x US-10-335-977-5526 (1-283)

QY 2 GCAGATTTTACAGAGCCAGATAGCCAAATGGGGTTTATTTTAAATTTGAGCATG 61  
Db 235 AlaGluPheTyArgGluProAspSerGlnLeuGlyValTyPheLeuAsnLeuSerMet 254  
QY 62 GGCAGATTTTAACTTTATTTAGTGTGTTTATTTAGGATTTTATGATGCTACA 121  
Db 255 GlyGlnIleLeuSerValPheMetValIleValSerLeuGlyIleLeuLeuTyzAlaThr 274  
QY 122 AAAAAATCTAAAAAATAAGGAA 145  
Db 275 LysAsnSerLysLysIleLysGlu 282

RESULT 10  
US-10-364-397-17  
Sequence 17, Application US/10364397  
Publication No. US20040014191A1  
GENERAL INFORMATION:  
APPLICANT: Minton, Nigel  
APPLICANT: Anlezark, Gill  
APPLICANT: Vaughan, Thomas  
TITLE OF INVENTION: Nitroreductase Enzymes  
FILE REFERENCE: 1581.0850001  
CURRENT APPLICATION NUMBER: US/10/364,397  
CURRENT FILING DATE: 2003-02-12  
PRIOR FILING DATE: 09/913,068  
PRIOR APPLICATION NUMBER: PCT/GB00/00431  
PRIOR FILING DATE: 2000-02-10  
PRIOR APPLICATION NUMBER: UK 9903019.9  
PRIOR FILING DATE: 1999-02-10  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 17  
LENGTH: 220  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-10-364-397-17

Alignment Scores:  
Pred. No.: 3.01e-15 Length: 220  
Score: 218.50 Matches: 59  
Percent Similarity: 48.89% Conservative: 51  
Best Local Similarity: 26.22% Mismatches: 94  
Query Match: 14.02% Indels: 21  
DB: 15 Gaps: 4



US-09-647-661-1 (1-900) x US-10-364-397-17 (1-220)

QY 151 ATGAATTTTGGATCAGAAAAGAGACAACTATTAAACGAGCGCCATTCTTGCAG 210  
Db 1 MetThrGlnLeuThrArgGlnGlnValLeuGluLeuPheHisGlnArgSerThrArg 20  
QY 211 ATGTTTGACGCCATTATAGTTTCTAGTGAAGAATTAGAAGAAATCGCTGAATCGCC 270  
Db 21 TyrTyrAspProThrLysLysIleSerAspGluAspPheGluCysIleLeuGluCysGly 40  
QY 271 AGGCTATCGCAAGCTCTTACACACGAGCGATGCGCATTTGTGTGATGTTACTAATAAG 330  
Db 41 ArgLeuSerProSerValGlySerGluProThrLysPheLeuValIleGlnAsnLys 60  
QY 331 GATTAAAAACAAATTCAGTCACAGCTACTTTAATGAAGAAATGATTAAAGCGCT 390  
Db 61 ThrLeuArgGluLysMetLysProPheSerTrpGlyMetIleAsnGlnLeuAspAsnCys 80  
QY 391 TCAGGTTAATGGGTATGCTCTTAAACAGCTAGCGAGTTGTTACACACGCGCATAC 450  
Db 81 SerHisLeuValValIleLeuAlaLysLys-----AsnAlaArgTyr 94  
QY 451 ATGCAAAACCTTTACCGGAGCTTTATAAGGTAGAGTGATCCCTTCTTTGCTCAATG 510  
Db 95 AspSerProPhePheValAspValMetAlaArgLysGlyLeuAsnAlaGluGlnGln 114  
QY 511 CTGGCGTGAGATTCACACACAGCATGCAAGATTAGAAAGCTATATTTAGAG----- 564  
Db 115 AlaAlaLeuThrLysTyrLysAlaLeuGlnGluAspMetLysLeuLeuGluAsnAsp 134  
QY 565 -----CAATGCTATATCGCTGTGGGCAATTTGCGATG 597  
Db 135 ArgThrLeuPheAspTrpCysSerLysGlnThrTyrIleAlaLeuAlaAsnMetLeuThr 154  
QY 598 GCGGTGAGCTTAATGGGATGGATAGTTGCATTATGAGGCTTTGATCCCTTTAAAGTG 657  
Db 155 GlyAlaSerAlaLeuGlyIleAspSerCysProIleGluGlyPheHisTyrAspLysMet 174  
QY 658 GGTGAAGTTTA--GAGAGCGTATCAATAGCTTAA-----ATCGCATGCTTG 705  
Db 175 AsnGluCysLeuAlaGluGluGlyLeuPheAspProGlnGluTyrAlaValSerValala 194  
QY 706 ATCGCTTTGGGCAAGAGGTGGCAGAGCGAGCGCAAAATCAAGAAATCAAAAGTTGAT 765  
Db 195 AlaThrPheGlyTyrArgSerArgAspIleAlaLysLysSerArgLysGlyLeuAspGlu 214  
QY 766 GCGATTACTTGTTG 780  
Db 215 ValValLysTrpVal 219

RESULT 11  
US-10-364-397-23  
; Sequence 23, Application US/10364397  
; Publication No. US20040014191A1  
; GENERAL INFORMATION:  
; APPLICANT: Minton, Nigel  
; APPLICANT: Anlezark, Gill  
; APPLICANT: Vaughan, Thomas  
; TITLE OF INVENTION: Nitroreductase Enzymes  
; FILE REFERENCE: 1581.0850001  
; CURRENT APPLICATION NUMBER: US/10/364,397  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR FILING DATE: 09/913,068  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: PCT/GB00/00431  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: UK 9903019.9  
; PRIOR FILING DATE: 1999-02-10  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 201  
; TYPE: PRT

; ORGANISM: Campylobacter jejuni  
US-10-364-397-23

Alignment Scores:  
Pred. No.: 58-13 Length: 201  
Score: 198.50 Matches: 52  
Percent Similarity: 50.26% Conservative: 46  
Best Local Similarity: 26.67% Mismatches: 68  
Query Match: 12.73% Indels: 29  
DB: 15 Gaps: 6

US-09-647-661-1 (1-900) x US-10-364-397-23 (1-201)

QY 169 GAAAAAGAGCAACTATTAAACGAGCGCCATTCTTCAAGATGTTTGAAGCCATTAT 228  
Db 2 LysLysGluLeuGluIlePheSerThrArgTyrSerCysArgAsnPheLysAsnGlu--- 20  
QY 229 GAGTTTTCTAGTGAAGAATTAGAGAAATCGCTGAATCGCCAGGCTATCGCAAGCTCT 288  
Db 21 LysLeuLysGluAspLeuAsnSerIleLeuGluIleAlaArgLeuSerProSerSer 40  
QY 289 TACAACACAGCAGCCATGTCATTTGTGATGTTTACTAATAGGATTTAAAAAACAATT 348  
Db 41 LeuGlyLeuGluProTrpLysPheIleValValGlnAspGluLysArgLysGluLeu 60  
QY 349 GCAAGTCACAGCTACTTTAATGAAGAAATG---ATTAAAGCGCTTCACGTTATGNG 405  
Db 61 Ser-----LysIleCysAsnGlnGlnLysHisValLysAspCysAlaAlaLeuIle 78  
QY 406 GTATGCTCT-----TTAAGACTAGCGAGTTG 432  
Db 79 IleIleSerArgLeuAspPheLeuAspTyrPheGluGluLysLeuArgLysArgAspMet 98  
QY 433 TTACACAGCGCCATTACATGCAAAACCTTTACCGGAGCTCTTATAGGTTAGAGTGATC 492  
Db 99 -----SerGluThrGluMetGlnLysArgLeu 107  
QY 493 CCTCTTTGCTCAATGCTTGGCGTGAGATTCAACACAGCATGCAAAAGATTAGAAAGC 552  
Db 108 AspThrTyrMetProPheLeuLys-----SerLeuAsnGlnGluGlnLysIleSer 124  
QY 553 TATATTTTGAAGCAATGCTATATCGCTGTGGGCGCAAAATTTGATGGCGGTAGCTTAATG 612  
Db 125 TyrAlaArgGluGlnAlaHisIleAlaLeuAlaSerIleLeuTyrSerAlaAsnAlaLeu 144  
QY 613 GGATTCGATGTTGCTATTTGAGGCTTTGATCCCTTTAAAGTGGGTCAAGTTTAGAA 672  
Db 145 AsnIleAlaSerCysThrIleGlyGlyPheAspLysGluLysLeuAspSerTyrLeuSer 164  
QY 673 GAGCGTATCAATAGCCTAAATCGCATGCTTATCGCTTTGGGC 717  
Db 165 LeuAspIleGlnLysGluArgSerSerLeuValValAlaLeuGly 179

RESULT 12  
US-10-335-977-6413  
; Sequence 6413, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-2114  
INFORMATION FOR SEQ ID NO: 6413:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 217 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCAT 1...217  
SEQUENCE DESCRIPTION: SEQ ID NO: 6413:  
US-10-335-977-6413

Alignment Scores:  
Pred. No.: 4,54e-12 Length: 217  
Score: 130.00 Matches: 54  
Percent Similarity: 51.63% Conservative: 57  
Best Local Similarity: 25.12% Mismatches: 86  
Query Match: 12.15% Indels: 18  
DB: 12 Gaps: 7

US-09-647-661-1 (1-900) x US-10-335-977-6413 (1-217)

QY 160 TTGGATCAGAAAAGAGACACATTTAAACGAGCGCATCTTTCAGAGATGTTGAC 219  
Db 1 MetAspArgGluGlnIleleAlaLeuGlnHisGlnArgPheAlaThrLysLysTyrAsp 20  
QY 220 AGCCATTATGAGTTTCTAGTGAAGAATAGAAATGCTGAAATCCGACGGCTATCG 279  
Db 21 ProAsnArgArgIleSerGluLysAspTrpGluValLeuValGluValGlyArgLeuAla 40  
QY 280 CCAAGCTTTACACACGAGCATGGCATTTTGTGATGTTACTAATAAGGATTAA 339  
Db 41 ProSerIleGlyLeuGluProIrpLysMetLeuLeuLysAsnGluArgMetLys 60  
QY 340 AAACAAATTCAGTCACAGCTACTTTAATGAAGAAATGATTAAAGCGCTTCAGCGTTA 399  
Db 61 GluAspLeuLysProMetAlaTrpGlyGlyLeuSerSerLeuGluGlyAlaSerHisPhe 80  
QY 400 ATGGTGGTATGCTTTAGACCTAGCGAGTTGTTACACAGCGCCATACATGCAAAAC 459  
Db 81 ValIleTyr-----LeuAlaArgLysGlyValThrTyrAspSerAspTyrValLysLys 98  
QY 460 CTTTACCGGAGCTTTATAAG-----GTTAGATGATC 492  
Db 99 ValMetHisGluValLysLysArgAspTyrAspThrHisSerArgPheAlaGlnIlele 118  
QY 493 CTTCTTTTGTCAAATGCTGGGTGAGATTCAACACAGCATGCAAGATTAGAAAGC 552  
Db 119 LysAsnPhe---GlnGluAsnAspIleLysLeuAsn---SerGluArgSerLeuPheAsp 136  
QY 553 TATATTTTAGACAATGCTATATCGCTGTGGGGCAAATTTGCTGCGGGGTGAGCTTAATG 612  
Db 137 TrpAlaSerLysGlnThrTyrIleGlnMetAlaAsnMetMetMetAlaAlaMetLeu 156  
QY 613 GGATTGGATGTTGCATTTATTCGAGCTTTGATCCCTTTAAAGAGGGGTGAAGTTTAA 672

Db 157 GlyIleAspSerCysProIleGluGlyTyrAspGlnGluLysValGluAlaTyrLeuGlu 176  
QY 673 GAGCGT-----ATCAATAAGCCTAAA-----ATCGCATGCTTGATCGCTTTGGCAAG 720  
Db 177 GluLysGlyTyrLeuAsnThrAlaGluPheGlyValSerValMetAlaSerPheGlyTyr 196  
QY 721 AGGGTGGCAGAGCAGCGACCCAAAATCAAGA---AAATCAAAAGTT 762  
Db 197 ArgAsnGlnGluIleThrProLysThrArgTyrLysThrGluVal 211  
RESULT 13  
US-10-364-397-12  
; Sequence 12, Application US/10364397  
; Publication No. US20040014191A1  
; GENERAL INFORMATION:  
; APPLICANT: Minton, Nigel  
; APPLICANT: Anlezark, Gill  
; APPLICANT: Vaughan, Thomas  
; TITLE OF INVENTION: Nitroreductase Enzymes  
; FILE REFERENCE: 1581.0850001  
; CURRENT APPLICATION NUMBER: US/10/364,397  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: 09/913,068  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: PCT/GB00/00431  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: UK 9903019.9  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 202  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-10-364-397-12  
Alignment Scores:  
Pred. No.: 1.41e-11 Length: 202  
Score: 185.50 Matches: 53  
Percent Similarity: 49.76% Conservative: 50  
Best Local Similarity: 25.60% Mismatches: 85  
Query Match: 11.90% Indels: 19  
DB: 15 Gaps: 7  
US-09-647-661-1 (1-900) x US-10-364-397-12 (1-202)

QY 184 CTATTAAACGAGCGCCATCTTTCAGAGATGTTTCAGAGCATATGATTTCTAGTGAA 243  
Db 7 ValLeuLysAlaArgAlaSerValLysGluTyrAspThrAsnAlaProIleSerLysGlu 26  
QY 244 GAATTAGAAAGAAATCGCTGAAATCGCAGGCTATCGCCAAAGCTCTTTACACACGCGCA 303  
Db 27 GluLeuThrGluLeuLeuAspLeuAlaThrLysAlaProSerAlaTrpAsnLeuGlnHis 46  
QY 304 TGGCATTTTGTGATGTTACTAATAAGGATTTAAATAAAACAATTCAGTCGACACTAC 363  
Db 47 TrpHisPheThrValPheHisSerAspGluSerLysAlaGluLeuLeuProValAlaTyr 66  
QY 364 TTTAATGAAGAAATGATTAAAGCGCTTCAGCGCTTAATGCTGATGCTTTTAAAGACCT 423  
Db 67 AsnGlnLysGlnIleValGluSerSerAlaValAlaIleLeuGlyAspLeuLysAla 86  
QY 424 AGCGAGTTGTTACACAGCGCCATTCATGCAAAACCTTTACCCGAG---TCCTTATAAG 480  
Db 87 AsnGlu-----AsnGly-----GluGluValTyrAlaGluLeuAlaSerGln 100  
QY 481 GTTAGAGTATGCTCTCTTTTGTCTCAATGCTTGGCGTGAGATTCACACACAGCATGCAA 540  
Db 101 GlyTyrIleThrAspGluLeuLysGlnThrLeuLeuGlyGlnIleAsnGlyAlaTyrGln 120  
QY 541 AGATTGAAGAGCTATATTTTAGAGCAATGCTAT-----ATCGCTGTGGGGCAA 588

Db 121 Ser---GluGlnPheAlaArgAspSerAlaPheLeuAsnAlaSerLeuAlaAlaMetGln 139  
QY 589 AITTCATGGCGTGAAGCTTAATGGGATTGGATAGTTCATTAATGGAGCGCTTGCATCCT 648  
Db 140 LeuMetIleAlaAlaLysAlaLysGlyTyrAspThrCysAlaIleGlyGlyPheAsnLys 159  
QY 649 -----TAAAGTGGCTGAAGTTTGAAGAGCGGTATCAATAAGCCTAAATCGCA 699  
Db 160 GluGlnPheGlnLysGlnPheAspIleSerGluA-gTyrValPro-----Val 175  
QY 700 TCGTTGATCGCTTTGGCGAAGGGTGGCAGAGCGAGCCAAATAATCAAGAAATCAAAA 759  
Db 176 MetLeuIleSerIleGlyLysAlaValLysProAlaHisGlnSerAsnArgLeuProLeu 195  
QY 760 GTTGATGCGATTACTTGGTTG 780  
Db 196 SerLysValSerThrTrpLeu 202  
RESULT 14  
US-10-169-048-2  
; Sequence 2, Application US/10169048  
; Publication No. US20030072769A1  
; GENERAL INFORMATION:  
; APPLICANT: Clarke, Edna Elizabeth  
; APPLICANT: Zhou, Liling  
; APPLICANT: Shea, Jacqueline Elizabeth  
; APPLICANT: Feldman, Robert Graham  
; APPLICANT: Holden, David William  
; TITLE OF INVENTION: Streptococcus Pyogenes Virulence Genes and Proteins And Their Use  
; FILE REFERENCE: GUE-97  
; CURRENT APPLICATION NUMBER: US/10/169,048  
; CURRENT FILING DATE: 2002-06-24  
; PRIOR FILING DATE: 2002-06-24  
; PRIOR FILING DATE: 2000-12-22  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 2  
; TYPE: PRT  
; LENGTH: 221  
; ORGANISM: Streptococcus pyogenes  
US-10-169-048-2  
Alignment Scores:  
Pred. No.: 3 57e-11 Length: 221  
Score: 182.00 Matches: 58  
Percent Similarity: 47.06% Conservative: 46  
Best Local Similarity: 26.24% Mismatches: 83  
Query Match: 11.67% Indels: 34  
DB: 14 Gaps: 6  
US-09-647-661-1 (1-900) x US-10-169-048-2 (1-221)  
QY 178 AGACAATATAACGAGCGCCATTCTTGCAGATGTTTGACAGCCATTATGAGTTTCT 237  
Db 10 GlnGlnAlaLeuHisPheArgThrAlaValArgValThrLysGluGlu---LysIleSer 28  
QY 238 AGTGAAATAGAGAAATCGCTGAATCGCAGGTATCGCAAGCTTACCAACAG 297  
Db 29 AspGluAspLeuAlaLeuIleLeuAspAlaAlaTrpLeuSerProSerSerIleGlyLeu 48  
QY 298 CAGCCATGCGATTGTTGATGGTTTACTATAAGGATTTAAAAAACAATAATGCGAGTGCAC 357  
Db 49 GluGlyTrpArgPheValValLeuAspAsnLysProIleLysGluGluLysProPhe 68  
QY 358 AGCTACTTTAATGAGAAATGATTAAAGCGCTTCAGCGTTATGTTGGTATGCTCTTTA 417  
Db 69 AlaTrpGlyAlaGlnTyrGlnLeuGluThrAlaSerHisPheIleLeuLeuIleAlaGlu 88  
QY 418 AGACTAGCGAGTTGTTTACCACACGCGCCATTATCATGCAAAACCTTTACCGGAGTCTTAT 477  
Db 89 Lys-----HisAlaArgTyrAspSerProAlaIleLysAsnSerLeu 102  
QY 478 AAGTTAGATGATCCCTCTTTTGTCTCAATGCTTGGCGTGAGATTCACACAGCATG 537

Db 103 LeuArgArgGlyIleLysGlu-----GlyAspGlyLeuAsnSerArgLeu 117  
QY 538 CAAAGATTGAAGAGCTATATTTAGAG----- 564  
Db 118 LysLeuTyrGluSerPheGlnLysGluAspMetAspMetAlaAspAsnProArgAlaLeu 137  
QY 565 -----CAATGCTATATCGCTGGGGCAAAATTTGCATGGCGGTGAGC 606  
Db 138 PheAspTrpThrAlaLysGlnThrTyrIleAlaLeuGlyAsnMetMetThrAlaAla 157  
QY 607 TTAATGGATTGATAGTTCATTTATGGAGGCTTTGATCCTTTAAAGTGGGTGAAGTT 666  
Db 158 LeuLeuGlyIleAspThrCysProIleGluGlyPheHisTyrAspLysValAsnHisIle 177  
QY 667 TTGAAGACAGCGT-----ATCAATAAGCTAAATCGCATGCTTGCATCGCTTTG 714  
Db 178 LeuAlaLysHisAsnValIleAspLeuGluLysGluGlyIleAlaSerMetLeuSerLeu 197  
QY 715 GCGAAGAGGGTGGCAGAGCGAGCCAA---AAATCAAGAAATAATCAAAAGTTGATGCGATT 771  
Db 198 GlyTyrArgLeuArgAspProLysHisAlaGlnValArgLysProLysGluGluValIle 217  
QY 772 ACT 774  
Db 218 Ser 218  
RESULT 15  
US-10-364-397-19  
; Sequence 19, Application US/10364397  
; Publication No. US2004001419A1  
; GENERAL INFORMATION:  
; APPLICANT: Minton, Nigel  
; APPLICANT: Anlezark, Gill  
; APPLICANT: Vaughan, Thomas  
; TITLE OF INVENTION: Nitroreductase Enzymes  
; FILE REFERENCE: 1581.0850001  
; CURRENT APPLICATION NUMBER: US/10/364,397  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: PCT/GB00/00431  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: UK 9903019.9  
; PRIOR FILING DATE: 1999-02-10  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 200  
; TYPE: PRT  
; ORGANISM: Synechocystis PCC6803  
US-10-364-397-19  
Alignment Scores:  
Pred. No.: 2 71e-10 Length: 200  
Score: 174.00 Matches: 53  
Percent Similarity: 47.50% Conservative: 42  
Best Local Similarity: 26.50% Mismatches: 87  
Query Match: 11.16% Indels: 18  
DB: 15 Gaps: 6  
US-09-647-661-1 (1-900) x US-10-364-397-19 (1-200)  
QY 193 GAGCGCCATTCTTGCAGATGTTTGACAGCCATTATGAGTTTCTAGTGAAGAATTGAA 252  
Db 9 GlnAlaArgSerValLysHisPheAspProAspHisArgLeuThrAlaGluGluArg 28  
QY 253 GAAATCGCTGAAATCGCCAGGCTATCGCAAGCTTACACACAGCAGCCATGCGATT 312  
Db 29 LysLeuHisGluAlaAlaIleGlnAlaProThrSerPheAsnIleGlnLeuTrpArgPhe 48  
QY 313 GTGATGGTTACTTAATAGGATTTAAAAAACAATAATGCGATGTCAGCATCTTTAATGAA 372

Wed Apr 21 10:19:02 2004

D <b>b</b>	49	LeuIleIlEAAspProGlnLeuArgGlnThrIleArg---GluLysTyrGlyAenGln	67
Q <b>y</b>	373	GAAATGATTAAAAGCGCTTCACGGTTAATGTGTGATGCTCTTTAAAGACCCTACGAGTTG	432
D <b>b</b>	68	AiaGlnWetThrAspAlaserLeuLeuIleValAlaLaaspValaenAlatPasp	87
Q <b>y</b>	433	TTACCACACGGCCATTACATGCACAACCTTTACCCGGAG-----TCTATAAGGTTAGA	486
D <b>b</b>	88	LysAspProAlaArgTyrTrpArgAsnAlaProArgGluValAlaAenTyrIleuValGly	107
Q <b>y</b>	487	GTGATCCCCTCTTTTGCTCAAAATGCTTGCGCGTAGATTCAACACACACAGCATCAAAAGATTA	546
D <b>b</b>	108	AlaIleAlaPheTyrGlyCly-----LysProGlnIleuGlnArgAsp	121
Q <b>y</b>	547	GAAGGTATATTTTTAGACAATGCTATATCCGTGTGGGGCAAATTTGCATGGCGGTGAC	606
D <b>b</b>	122	GluAlaGln-----ArgSerileGlyMetAlaMetGlnAenLeuMetIleuAlaLaLys	139
Q <b>y</b>	607	TTAATCGGATTGGATAGTTCATATTGGAGCGCTTGATCCTTTAAAGTGGGTGAGTT	666
D <b>b</b>	140	AlaMeEGlyTyRAspSerCysPrometIleGlyPheaSpleuGlnLysValAlaGluLeu	159
Q <b>y</b>	667	TTAAGACAGCGTATCAATAAAGCCT-----AAAAATCGCATCCTGTATCGCTTTTGGCG	717
D <b>b</b>	160	-----ValLysLeuProAlaAspTyrAlaIleGlyPrometValAlaIleGly	175
Q <b>y</b>	718	ACAGCGGTGGCAGACCGGCCCAAAAATCAAGAAAAATCAAAAGTTGATGCGATTACTTGG	777
D <b>b</b>	176	LysEArThrGluAspAlaArgAlaLyscylGlyGlnThrProLeuGluGluLeuValtrp	195

Search completed: April 20, 2004, 19:22:36  
Job time : 68 secs